

Projector-Based Illumination and Display Techniques

by

Oliver Bimber
Bauhaus-University Weimar

Submitted to the department of Computer Science
in partial fulfillment of the

Habilitation

procedure at the

Technical University of Munich

To Mom

1 Introduction

Rapid advances in electro-mechanics and optics have increased the capabilities of projectors in terms of spatial resolution, brightness, dynamic range, throw-ratio, and speed. Cost reductions, availability, and the fact that projectors (in contrast to flat panels) can display images that are much larger than the devices themselves made them a mass-market product. Besides for (home-)entertainment, education and business purposes, high resolution tiled screens and immersive surround screen projections are used for visualizations of scientific, engineering and other content. Emissive CRT technology is losing more and more ground to light-valve technology, like LCD, LCOS and especially DLP.

Compact pocket projectors that are running from battery and communicate to laptops or cell phones via WiFi will support a maximum level of mobility in future. Besides portable projectors, projection systems will be integrated into next-generation mobile devices, such as cell-phones, PDAs and digital cameras to support built-in flat-panel screens. There is no doubt that LED technology will become bright enough to keep up with today's projector lamps. Yet one question still remains: What to project on if not carrying around a projection canvas?

Other applications do not require mobility and rather benefit from high quality spatial presentations. Examples range from edutainment in museums (such as storytelling projections onto natural stone walls in historical buildings) over architectural applications (such as augmentations of complex illumination or material simulations in real building structures) to video presentations during live stage performances.

For all of these cases, real-time image correction techniques are required that enable presentations of visual content on surfaces that are not optimized for projections. Geometric and radiometric distortions, local and global illumination effects, and defocus cause the main visual artifacts that have to be neutralized when approaching a result that comes closer to projections onto conventional (e.g., white planar) screens.

Physically, projectors represent point light sources. Their capability of spatially modulating luminance and chrominance on a per-pixel basis, however, allows for computing and creating almost arbitrary shading effects synthetically. In contrast to seeing projectors as mobile or spatial displays, they can be used as controllable light sources for synthesizing a virtual illumination in a real environment.

The challenge of a projector-based illumination is to produce a defined lighting situation without the necessary light sources. The only available light sources are the projectors themselves. Consequently, illumination images have to be computed for each projector with the following objectives: First, they must neutralize the physical illumination effects that are caused by each projector as a real point-light source. Second, they have to produce the defined virtual lighting situation synthetically.

Applications of projector-based illumination exist in areas, such as physical re-illumination of real artifacts in a computational photography or a live-presentation context, replaying optical holograms with digital light and -potentially- illuminating future television studios.

The overall focus of the work that is summarized in this document is on real-time computer graphics methods for computing projection images that produce the desired effects on complex surfaces and optics – either for illumination or for display purposes. Thereby, a broad spectrum of problems and applications is addressed. While section 2 gives a chronological summary of research results and funding sources, section 3 presents details on contributions and differentiations from related work in a more theme oriented way. Finally, section 4 gives an outlook on future directions. The relevant publications are attached to this summary.

2 Summary of Results

The research focus throughout the Ph.D. work (2000-2002) was on *spatial optical see-through displays*. These are augmented reality displays that apply spatial optical elements to overlay computer graphics optically over real objects. One result is the Virtual Showcase [Bim01b], which is a showcase-like museum display that is capable of presenting real artifacts augmented with interactive graphical elements to multiple museum visitors simultaneously. The goal of the Ph.D. work was to develop new real-time rendering techniques that allow integrating optical elements (such as curved or planar mirrors and lenses) into the optical path of an information display [Bim00, Bim01a, Bim01b, Bim03a]. Several problems that are related to optical see-through displays, however, remained unaddressed. Examples are inconsistent occlusion and illumination effects between real and virtual objects when optical overlays are produced.

Projectors can act as light-sources that allow illuminating real objects with digital light – thus giving full spatial and temporal control over physical shading and shadows. Consequently, after the Ph.D. work (2002-present) a main research focus was on novel projection techniques. Projector-based illumination together with spatial optical see-through, for instance, proved to provide realistic augmentations.

One first projector-based illumination technique for creating *correct occlusion effects* for optical see-through setups was presented in 2002 [Bim02]. View-dependent occlusion shadows were projected onto the real surfaces that are located behind virtual objects. This resulted in a correct occlusion of real objects by virtual ones. This approach was implemented and tested in the context of the Virtual Showcase display. Hardware extensions for projecting light into the showcase and rendering techniques for displaying occlusion shadows for single and multi-user scenarios as well as for single and multi-light-projector configurations were described and presented.

Further techniques that create a *consistent illumination* between real and virtual objects were then presented in 2003 [Bim03b]. Projectors and cameras were applied to capture reflectance information from diffuse real objects and to illuminate them under new synthetic lighting conditions. Matching direct and indirect lighting effects, such as shading, shadows, reflections and color bleeding can be approximated at interactive rates in such a controlled mixed environment.

Holography and computer graphics are being used as tools to solve individual research, engineering, and presentation problems within several domains. Up until today, however, these tools have been applied separately. The intention of the work that was carried out between 2003 and 2005 was to combine both technologies to create a powerful tool for science, industry and education. Therefore, the possibility of integrating computer generated graphics into optical holograms has been investigated. Projectors play a key-role for making this possible: The main advantage for using projectors instead of analogue light bulbs for replaying optical holograms is that the reference beam produced by a projector that is used to reconstruct the hologram can be digitized. Thus it is possible to control the amplitude and wavelength of each discrete portion of the reference beam over time. Consequently, projector-based illumination can be applied that –using the holographic film itself as an optical combiner– allows *combining holographic and graphical content in a consistent way* [Bim04a, Bim04b, Bim05c, Bim06d, Bim06e].

In contrast to projector-based illumination, projecting images that contain a pictorial content onto everyday surfaces causes visual artifacts. Reflected images are geometry distorted, color

blended and partially defocused. Since 2004-present, research has also focused on projector-camera techniques that compensate these distortions.

Multi-projector capable radiometric compensation techniques were developed from 2003-2005. Thereby, multiple projectors overcome the intensity limitations of single projector approaches. The first prototype (2003) performed a **real-time color correction to augment two-dimensional surfaces**, such as paintings, with images, videos or interactive elements [Bim05a]. An adapted version of this technique (2004) was then used for displaying images, videos and interactive content on **geometrically complex surfaces**, such as window curtains, stone walls and wall-papered walls, etc. [Bim05b]. The projected images were geometry and color corrected in real-time and on a per-pixel basis and appeared undistorted from a single sweet-spot.

In 2004/2005 a **view-dependent image-based and geometric warping technique** was presented that enabled a radiometrically compensated stereoscopic visualization, on ordinary (geometrically complex, colored and textured) surfaces within everyday environments [Bim05e]. This technique was based on unstructured Lumigraph rendering.

The radiometric compensation techniques described above address only local illumination effects. Global effects that result from an indirect illumination (e.g., scattering of light from surface to surface) have to be considered as well. Therefore, a **real-time reverse radiosity** method for compensating indirect scattering effects was developed in 2005. It computes a **numerical solution** [Bim06a] directly on the GPU and is implemented with pixel shading and multi-pass rendering, which together realize a Jacobi solver for sparse matrix linear equation systems. Later in 2005, an **analytical solution** [Bim06b] to this problem was found. This method was validated and evaluated based on a two-sided projection screen. The same rendering technique, however, is applicable to arbitrary surfaces.

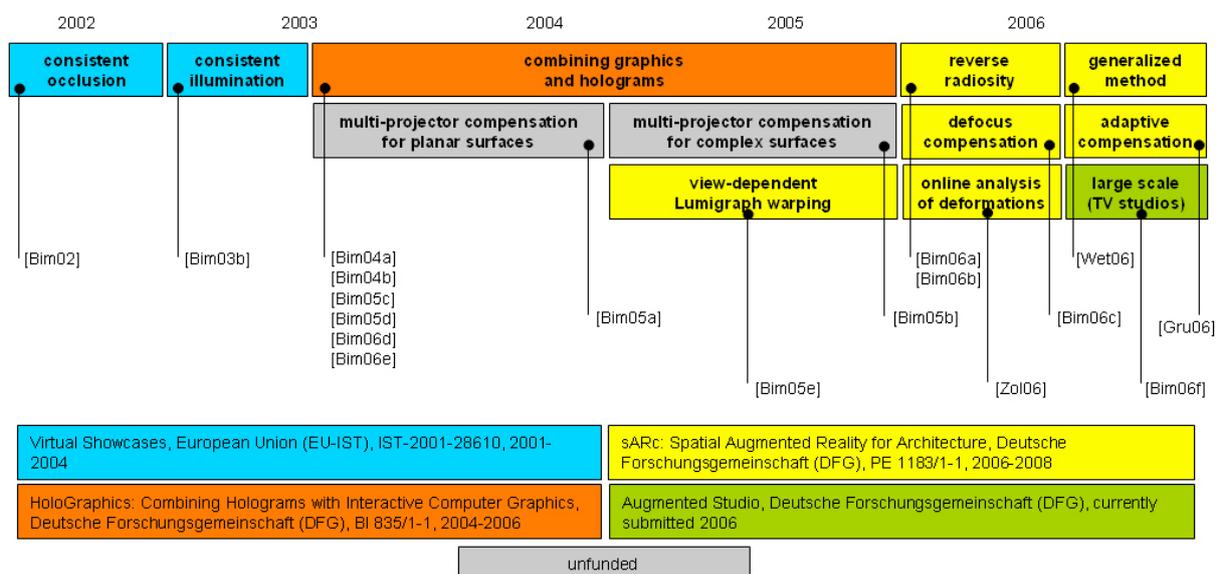
So far, the developed compensation techniques correct colors, intensities and positions of pixels to achieve an overall image consistency. However, one important pixel property remains unaddressed so far: its focus. Images that are projected onto non-flat surfaces are partially blurred. Important features that are required for disparity-based depth perception are washed out. Stereoscopic images become more difficult or even impossible to fuse. To solve this problem, a novel **multi-focal projection** concept has been developed in 2005 that applies conventional projectors and camera feedback [Bim06c]. Multiple projectors with differently adjusted focal planes, but overlapping image areas are used. They can be either positioned unstructured in the environment, or can be integrated into a single projection unit. The defocus created on an arbitrary surface is estimated automatically for each projector pixel. If this is known, a final image with minimal defocus can be composed in real-time from individual pixel contributions of all projectors. This technique is independent of the surfaces' geometry, color and texture, the environment light, as well as of the projectors' position, orientation, luminance and chrominance.

As an extension to the view-dependent image warping techniques that were developed in 2004/2005, a new hybrid technique for correcting geometric distortions was developed in 2005/2006. Rather than an offline calibration for multiple source perspectives, it supports an **online analysis of the image distortions** from a new target perspective [Zol06]. It investigates the optical flow that results from perspective distortions during motions of the observer and tries to use this information for computing the correct image warping. If this fails due to an unreliable optical flow analysis, an accurate –but slower and perceivable– structured light projection is automatically triggered.

In 2006 a **generalized method** was found that **unifies all previously developed image correction techniques** [Wet06]. Acquiring the forward light transport between a projector and a camera allows capturing the entire light modulation of a projected pattern within an arbitrarily complex scene - including all local and global illumination effects. This represents a 4D slice of the 8D reflectance field and can be described as a linear equation system.

Replacing the camera image in the forward light transport equation with a desired picture enables to perform an image compensation that corrects geometry, radiometry, local and global effects as well as defocus by solving this equation system for the corresponding projector pattern.

As a result of a radiometric compensation, the brightness and the contrast of the input image is reduced compared to a conventional projection onto a white canvas. If the input image is not manipulated in its intensities, the compensation image can contain values that are outside the dynamic range of the projector. They will lead to clipping errors and to visible artifacts on the surface. In 2006, a novel algorithm was presented that dynamically adjusts the content of the input images before radiometric compensation is carried out [Gru06]. This reduces the perceived visual artifacts while simultaneously preserving a maximum of luminance and contrast. This *adaptive compensation algorithm* takes the capabilities of the human visual system into account and is the *first of its kind to run in real-time*.



Overview over research results, publications and funding.

Virtual studio technology plays an important role for modern television productions. Blue-screen matting is a common technique for integrating real actors or moderators into computer generated sceneries. Augmented reality offers the possibility to mix real and virtual in a more general context. In 2006, a new technological approach for combining real studio content with computer-generated information was proposed [Bim06f]. Digital light projection allows a controlled spatial, temporal, chrominance and luminance modulation of illumination. This opens new possibilities for TV studios, such as dynamic re-illumination with integrated imperceptible coded patterns that support continuous online-calibration, in-shot camera tracking, and the acquisition of scene depth. The development of new projection techniques that support *large-scale* applications (especially in the *context of modern television studios*) belongs to the current research activities.

3 Contribution Details

This section describes the contributions that have been made in the context of this habilitation and differentiates them from related and previous work. After a short introduction to each topic, the related work is discussed and the contributions made are outlined and highlighted in *bold-italic*.

3.1 Projector-Based Illumination Techniques

Projectors are point-light sources with the capability of modulating light spatially and temporally. A combination of projector and camera represents a scanning light source that is capable of analyzing the surrounding environment and of producing artificial lighting conditions by computing and projecting appropriate intensity images. Multiple networked projectors can synthetically illuminate large environments and dynamically respond to changes.

The concept of such two-way optical transducers –called “I/O bulbs”– that consist of pairs of projectors and cameras has been described by Underkoffler, et al. [Und99]. Based on this early suggestion, several research activities were carried out under the notation of “projector-based illumination”.

The following subsections summarize the contributions that have been made to the field of projector-based illumination: *Visualization problems of spatial optical see-through displays have been solved* and *interactive computer graphics could be combined with optical holograms* using *new projector-based illumination techniques*.

3.1.1 Consistent Occlusion for Spatial Optical-See-Through Displays

Spatial optical see-through displays, such as the Virtual Showcase [Bim01b], have several advantages over mobile displays. They provide high resolution graphics, improved consistency of eye accommodation and convergence, little motion sickness potential, and the possibility of integration into common everyday environments. One of the main challenges for spatial as well as for mobile optical see-through systems is the generation of correct occlusion effects between virtual and real objects. Light that is reflected from the surfaces of real objects interferes with the light that is redirected by an optical combiner (e.g., a half-silvered mirror) used by the display. Consequently, graphical overlays appear semitransparent. Additionally, shadows of virtual objects cast onto real ones and consistent illumination of the real and virtual scenery are often difficult to achieve.

Kiyokawa et al. [Kiy00] presented ELMO, an optical see-through head-mounted display that supports mutual occlusion. ELMO uses half-silvered mirrors as optical combiners and an additional semi-transparent LCD panel in front of the conventional optics. The LCD panel is used to selectively block the incoming light on a per-pixel basis. This enables virtual objects to occlude real ones. A head-attached depth sensor allows them to acquire depth maps of the real environment in real time. This makes the occlusion of virtual objects by real ones possible. ELMO faces a number of problems that are linked to the LCD panel: light attenuation caused by the LCD panel, and low response time and resolution of the LCD panel. However, as the first functioning system of its kind, it effectively addresses the occlusion problem of optical see-through head-mounted displays.

Noda et al. [Nod99] presented a stationary optical see-through display that uses a projector to illuminate real objects selectively – not lighting those areas that are overlaid by graphics.

Noda's system is strictly limited in several points: First, a dark surrounding environment is required, which constrains the possible applications. Second, view-dependent rendering is not possible. The observer's viewpoint has to match with the center of projection of the projector since the illumination pattern is rendered from this point using a normal on-axis projection. In this special case no depth information of the real environment is required for a correct rendering. Last, stereoscopic rendering is not provided.

Naemura et al. [Nae02] proposes an approach that is technically similar to Noda's. The conceptual difference, however, is that a hand-held projector is applied as a real flashlight to interactively generate shadow effects of virtual objects on real surfaces. This approach does not address the occlusion problem of optical see-through displays, but focuses on enhancing interactive mixed reality applications by providing additional visual cues through shadows. As in Noda's case no depth information of real objects is needed.

A new projector-based illumination method for creating consistent occlusion effects for spatial optical see-through displays was introduced in [Bim02]. It is similar to Noda's approach [Nod99]. However, it *supports view-dependent and stereoscopic rendering for single and multiple users* and *does not require the illumination being projected from the user's point of view*. Depth knowledge of the real scenery to support both is required—the occlusion of real objects by virtual ones and vice versa. Using calibrated projectors, shadows are dynamically projected on real objects' surfaces wherever graphics is overlaid. These shadows are not directly visible to the observers, since they are purposely occluded by the overlaid graphics. They are called “occlusion shadows”. Phantom bodies that represent real objects are rendered into the depth buffer to occlude virtual objects behind them. The combination of occlusion shadows and depth buffering effectively solves the occlusion problem for spatial optical see-through displays, such as the Virtual Showcase. Being implemented in two rendering passes directly on the graphics card allows creating consistent occlusion effects in real-time.

A main drawback of this approach is its *limited support for multiple users*: If the same real surface is simultaneously visible from multiple points of view (as it is the case for different observers), individual occlusion shadows that project onto this area is also visible from different viewpoints at the same time. Knowing the reflectance information of the surface, however, leads to a *general solution* [Bim03b]: In addition to the virtual scene, the portions of the real scene (i.e., its captured and registered reflectance map) that is covered by the occlusion shadows of all other observers are rendered as optical overlays. This creates seamless transitions between the real and the virtual parts, and provides consistent occlusion effects for multiple users.

3.1.2 Consistent Illumination for Spatial Optical-See-Through Displays

Besides consistent occlusion effects, a consistent lighting situation between real and virtual environments might be important for several convincing augmented reality applications. Problematic is the fact that real objects are illuminated by physical light sources while the shading and shadowing for virtual objects are computed based on defined virtual light sources. Since the two lighting conditions do not necessarily have to match, a well perceivable inconsistency between the illumination of real and virtual objects is the result.

A method for creating a consistent illumination between real and virtual components within a spatial optical see-through display—such as the Virtual Showcase—was presented in [Bim03b]. Combinations of projectors and cameras are applied to capture reflectance information from diffuse real objects and to illuminate them under new synthetic lighting

conditions. To *handle global lighting effects*, such as diffuse interreflections, an offline radiosity procedure is outlined that consists of multiple rendering passes. This method supports only static scenes, but can be used in combination with head-tracking. For *supporting local lighting effects* (such as simple shading, hard shadows and simple reflections) hardware accelerated techniques are described which allow to achieve interactive frame rates for dynamic scenes.

Consistent and matching shading effects on real surfaces from virtual light sources can be achieved by using projectors that display appropriate irradiance maps on the real objects. Raskar et al. [Ras01] show how to compute an irradiance map to lift the radiance properties of neutral diffuse objects with uniform white surfaces into a pre-computed radiance map of a virtual scene illuminated by virtual light sources. This technique can be used to augment diffuse, white objects with textures that can contain synthetic illumination effects. The method presented in [Bim03b], however, supports *the physical re-illumination of real (i.e., arbitrarily textured and colored – but diffuse) objects based on a defined (virtual) lighting situation*.

A two-pass rendering algorithm first simulates the lighting that is created on the real surfaces by all projectors as physical point light sources. In a second pass, the synthetic illumination created from virtual light sources is computed. A final step processes both simulations in such a way that physical lighting effects are neutralized, while the virtual illumination is synthesized on the real surfaces. Furthermore, hardware-accelerated shadow and cube mapping techniques are used to cast hard shadows from virtual objects onto real ones and vice versa, and to create simple direct reflections of real objects in virtual ones. All steps are carried out on the graphics card and support real-time frame rates. They lead to consistent illumination effects on real and virtual objects.

3.1.3 Replaying Optical Holograms with Digital Light

Many areas in science, entertainment, education, and engineering would be unimaginable without the aid of 2D or 3D computer graphics. The reason for the success story of computer graphics might be its interactivity, which is an important property that is still not provided efficiently by competing technologies – such as holography.

While display holography is limited to presenting a non-interactive content, electroholography or computer generated holograms (CGH) facilitate the computer-based generation and presentation of holograms at interactive rates [Kol89, Luc97, Sli04, Sli05]. Holographic fringes can be computed by either rendering multiple perspective images, then combining them into a stereogram [Luc95], or by simulating the optical interference and calculating the interference pattern [Luc93]. Special display systems dynamically generate the output wavefront from the computed fringe data. Since creating an electrohologram requires processing, transmitting, and storing a massive amount of data, today's computer technology still sets the limits for this technology. To overcome some of these performance and storage issues, advanced reduction and compression methods have been developed that create truly interactive electroholograms. Unfortunately, most of these holograms are relatively small, low resolution, and cover only a marginal color spectrum. However, recent advances in consumer graphics hardware may reveal potential acceleration possibilities that can overcome these limitations [Pet03].

Especially display holography has conquered several public application domains, such as museums, theme parks and trade shows. Displaying artifacts virtually removes the need to build physical replicas or showcase originals. This can save display space. But the true reason for displaying holograms might mainly lie in the fascination of viewing three-dimensional images. In addition, holograms can be used to make engineering, medical, dental,

archaeological, and other recordings—for teaching, training, experimentation and documentation. Archaeologists, for example, use optical holograms to archive and investigate ancient artifacts [Dre97,Dre00].

In contrast to most stereoscopic or autostereoscopic graphics displays, holographic images can provide all depth cues—perspective, binocular disparity, motion parallax, convergence, and accommodation—and theoretically can be viewed simultaneously from an unlimited number of positions. Today, computer graphics and raster displays offer a megapixel resolution and the interactive rendering of megabytes of data. Optical holograms, however, provide a terapixel resolution and are able to present an information content in the range of terabytes in real-time. Both are dimensions that will not be reached by computer graphics and conventional displays within the next years.

Obviously, one has to make a decision between interactivity and quality when choosing a display technology for a particular application. While some applications require high visual realism and real-time presentation (that cannot be provided by computer graphics), others depend on interactivity (which is not possible with analog display holograms).

Several possibilities have been investigated that allow *merging computer generated graphics and white-light optical holograms* [Bim04a, Bim04b, Bim05c, Bim06d, Bim06e]. The goal is to combine the advantages of conventional holograms (i.e. extremely high visual quality and realism, support for all depth queues and for multiple observers at no computational cost, space efficiency, etc.) with the advantages of today's computer graphics capabilities (i.e. interactivity, real-time rendering, simulation and animation, stereoscopic and autostereoscopic presentation, etc.).

As mentioned before, conventional projectors represent point sources. They are well suited for reconstructing most white-light reflection and transmission holograms. The main advantage for using projectors instead of analog light bulbs -like halogen spots- to replay a hologram is that the light frustum produced by a projector can be dynamically digitized. Thus it is possible to control the intensity and color of each discrete portion of the reference beam over time.

Early experiments with projectors for reconstructing optical holograms have been made in the art and engineering domains. In some art installations, optical holograms have been linked with time-based media, such as slides, film-loops or color patterns that are projected onto them to achieve artistic effects [Vil93, Vil94]. Others have redirected projected light with multiple mirrors to simulate different light sources. The goal was to achieve dynamic fluctuation effects with optical holograms [Oka97, Yam98].

A hybrid display approach has already been described earlier that combined a transmission hologram with a liquid crystal display to realize a new user interface for business machines, such as photocopiers [And89]. An analog point light source was used to illuminate a transmission hologram which was mounted behind an LCD panel. In this case it was not possible to control the reconstruction of the holographic image at discrete areas.

Using projected light as a reference beam it is possible to reconstruct a hologram only partially, leaving gaps where graphical elements can be inserted. Interactive computer graphics can then be displayed behind the holographic film and can be viewed through these gaps simultaneously with the holographic content.

The replayed hologram's intensity is proportional to the intensity of the projected reference beam. In addition to using an incomplete illumination for replaying a fraction of the hologram, intensity variations of the projected light permit local modification of its brightness (i.e., the local brightness of the replayed image beam). Thereby, rendering and illumination are view-dependent and have to be synchronized. Projector-based illumination techniques –similar to the ones described for real objects– can be used for *creating consistent occlusion and illumination effects between graphical and holographic content*. This novel *key concept*,

real-time rendering techniques, and proof of concept examples have been described for the first time in [Bim04a]. *Display prototypes, rendering and interaction techniques* for a variety of different types of *small scale holograms* [Bim05d] and for *large scale holograms* [Bim06d, Bim06e] have been developed and realized.

Since depth information of graphical and holographic content is required to support the corresponding rendering techniques, depth acquisition from holograms represents a main challenge. Having depth information of both –holographic and graphical content– allows computing a binary or shaded illumination image which replays the hologram in such a way that graphical components can be integrated consistently. This includes correct occlusion and shading effects. A *mathematical framework* has been introduced *for reconstructing depth information from small scale holograms with flatbed-scanners* [Bim05d]. Furthermore it was shown how to extract *depth from large scale holograms with multi-perspective range scans* [Bim06d, Bim06e].

3.2 Projection-Based Display Techniques for Complex Surfaces

The main difference between projector-based illumination and projector-based display techniques is that images generated for the latter carry a pictorial content rather than a pure illumination.

The ability to generate images that are larger than the actual display device virtually anywhere is an interesting feature for many applications that cannot be provided by desktop screens. Several research groups exploit this potential by using projectors in unconventional ways to develop new and innovative information displays that go beyond simple screen presentations. The Luminous Room [Und99] for instance, described an early concept for providing graphical display and interaction at each of an interior architecture space’s surface. With the Everywhere Displays projector [Pin01], this concept has been extended technically by allowing a steerable projection using a pan/tilt mirror. A similar approach was followed by Ehnes, et al. [Ehn04]. It was also demonstrated how context-aware hand-held projectors –so-called iLamps– can be used as mobile information displays and interaction devices [Ras03]. These are only a few representative examples of approaches that augment pictorial information on everyday surfaces that are not optimized for a projection. Except simple geometries, most of the remaining surface properties are not taken into account by these techniques. This leads to image distortions that can actually be compensated with more advanced methods.

The subsections below summarize the contributions that have been made to the field of *image compensation techniques for projecting onto complex surfaces*: radiometric compensation to *reduce chrominance and luminance distortions*, geometric image warping to *support view-dependent applications*, defocus compensations to *increase focal depth*, and *compensating global lighting effects*.

3.2.1 Radiometric Compensation

Projector-camera systems have been used together with radiometric compensation algorithms to project onto complex everyday surfaces, like papered walls or structured table tops. They measure the surface reflectance as well as the contribution of the environment light by applying structured light projection and camera feed-back. These parameters are then used for correcting the projected images in such a way that blending artifacts with the underlying surface are minimized.

Nayar et al. [Nay03], for instance, expressed the color transform between each camera and projector pixel as pixel-individual 3x3 color mixing matrices. These matrices were estimated from measured camera responses of multiple projected sample images. They can be continuously refined over a closed feedback loop and were used to correct each pixel during runtime. Later, a refined version of this technique was used for controlling the appearance of simple two- and three-dimensional objects, such as posters, boxes and spheres [Gros04]. Fujii et al. [Fuj05] presented a method that utilizes a co-axial alignment of projector and camera for dynamic compensation on non-static surfaces. A closed feedback loop was used in this case to re-adjust the compensations over time while either surface or projector-camera pair can be moved.

Problematic for these methods is the fact that they will fail if the underlying pigments absorb too much of the projected light, and the remaining reflected portion is too low to approximate the desired image well enough. This problem is known under the term “clipping”. The contributions in the area of basic radiometric compensation are two methods that minimize such clipping errors.

To solve this problem, the first method [Bim05a] applied a new film material that has two properties: first, it is completely transparent and second, it diffuses a fraction of the light projected onto it. The film consists of an even deposition of fine particles on both sides of a polyester base with no visible artifacts. Planar surfaces, such as paintings or other pictorial artworks can be covered with *the transparent film material*. A *new mathematical framework and a real-time rendering algorithm* have been developed to support the transmission and reflection behavior of the film for radiometric compensation. Clipping is reduced by the portion of light that is directly being reflected, while the transmitted portion has to be compensated because it is blended with the underlying pigments. Yoshida et al. [Yos03] described a virtual retouching technique that applies projectors for reconstructing and enhancing the faded colors of paintings by projecting light directly onto the canvas. An affine correlation between projection and the result captured by a camera was established manually for each pixel. Users can then retouch the original painting interactively via a desktop GUI. Clipping artifacts that are due to the limited brightness and dynamic range of the projector, however, are not reduced by this method.

The second method is *the first radiometric compensation algorithm that supports multiple interplaying projectors* [Bim05b]. While the method described above is applicable only to simple (mainly planar) surfaces, this approach can be applied to arbitrarily complex surfaces. To achieve real-time frame rates, per-pixel displacement mapping for geometric warping as well as per-pixel radiometric compensation are carried out entirely on the GPU. This represents one of the *first real-time capable radiometric compensation algorithms capable of compensating dynamic content*, such as videos. Approximately 100fps could be achieved for compensating a PAL resolution video on the GPU on off-the-shelf graphics cards. In contrast to existing methods it uses single disjoint camera measurements of surface reflectance, environment light contribution and projector form-factor components for radiometric compensation using hardware accelerated pixel shaders. Clipping errors are reduced by this method, since it allows a manifold of light to arrive at the surface – resulting from individual contributions of multiple projectors. Shadow casts produced by individual projectors can also be neutralized by multiple interplaying projectors. However, it does not take the color mixing of the individual RGB channels into account, as in [Nay03].

Recent algorithms extend basic radiometric compensation by first varying the input image to achieve an optimized compensation quality with minimized clipping artifacts. Wang et al.

[Wan05] presented the first technique that scales the overall intensity of the input image until clipping errors that result from radiometric compensation are below a perceivable threshold. Their computational intensive numerical minimization can only be applied to static monochrome images and surfaces.

Park et al. [Par06] described a technique for increasing the contrast in the compensation image by applying histogram equalization to the colored input image. This does not preserve the contrast ratio of the original input. Consequently, the image content is modified significantly. The problem of occurring clipping errors is not considered by this method.

A complex framework for computing an optimized photometric compensation was presented by Ashdown et al. [Ash06]. The surface's reflectance is scanned with a color calibrated HDR camera. The captured data and the image content are then transformed into the device independent CIE L^*u^*v color space in which color distances are based on the human visual perception. The chrominance values are fitted into the gamut of each projector pixel. Finally a luminance fitting is applied with a relaxation method based on differential equations. The compensation algorithm presented in [Nay03] is then used with the adjusted image rather than with the original image. This method can achieve optimized results even for projections onto surfaces with extremely varying reflectance properties or high saturation - but for static images only.

All of these techniques lead to reduced clipping artifacts and consequently to an increased visual quality compared to basic compensation methods that do not pre-adapt the input images. However, due to their computational complexity that can mainly be contributed to numerical minimizations, a real-time compensation cannot be achieved. This constrains them to the presentation of still images. Animated content, such as movies, can only be displayed after a time consuming pre-correction. This, however, is impractical for most applications. It is particularly useless if surface and setup do not remain completely static, such as it is the case for mobile projectors that will require a flexible and frequent re-calibration on different surfaces. Furthermore, distributed content, such as DVDs or broadcasted media cannot be pre-corrected for multiple individual surfaces. Finally, it is clear that real-time dynamic content, such as interactive applications cannot be presented at all.

A *novel algorithm* has been developed that dynamically adjusts the content of the input images before radiometric compensation is carried out by considering the capabilities of the human visual system [Gru06]. This *reduces the perceived visual artifacts while simultaneously preserving a maximum of luminance and contrast*. The *adaptation algorithm* is implemented entirely on the GPU and is *the first of its kind to run in real-time* (~35fps for PAL resolution videos on off-the-shelf graphics cards).

The algorithm applies the basic compensation scheme presented in [Bim05b], but only minor modifications to the adaptation algorithm are necessary to use other techniques instead. It analyzes the input image as well as the projection surface. The results are used for a global intensity adaptation of the input image. The scaled image is compensated and analyzed for clipping errors. Depending on these errors, the input image is globally re-scaled and locally adapted. The final result is compensated again and projected. All adaptation parameters are temporally adjusted to ensure smooth intensity transitions and to avoid visible flickering. The objective enhancement of the perceived visual quality for projected animated content was confirmed by an informal user study. Such a technique will be essential to support the portability of pocket projectors and projection systems that are integrated into mobile devices, such as cell-phones, PDAs and digital cameras.

3.2.2 View-Dependent Geometric Warping

For projecting multiple undistorted images registered onto planar surfaces the estimation of projector-to-camera homographies is very common (e.g., [Yan01a, Che02]). For geometrically non-trivial surfaces (i.e., uniformly colored surfaces with known geometry), the extrinsic and intrinsic parameters of each projector can be calibrated relative to the surface. Projective texture-mapping can then be applied for warping images from the perspective of a head-tracked observer into the perspectives of each projector. This supports a view-dependent warping of perspective images that are projected onto geometrically non-trivial surfaces [Ras99].

Projective texture mapping is based on a simple pinhole camera model and can consequently not handle non-linear effects, such as radial distortion caused by projector lenses. This produces slight misregistrations on the surface in the order of several pixels.

A precise projection of pixels onto corresponding surface pigment is extremely important – especially for noisy surface textures. Even small misregistrations can cause profound visual color and brightness artifacts, such as bright or dark spots that highlight the real surface features. The human visual system is extremely sensitive to such artifacts, and depth perception of virtual content is disturbed by highlighted real surface features. Note that geometric image distortions which result from misregistrations are not as critical.

To achieve an adequate registration of projector pixels on textured surfaces, 2D look-up tables that map every pixel from camera space to projector space and vice versa can be used. This is more precise than the determination of the mapping analytically or algorithmically. Projected structured light patterns are applied for measuring an unambiguous mapping. All radiometric parameters (surface reflectance, environment light contribution and projectors' form factor components) are measured also from the perspective of the camera. This ensures a direct look-up of all stored values in a fragment shader for each projector pixel. Since all parameters are measured for the perspective of the calibration camera, a geometric and radiometric compensation is only possible for this single perspective.

Inspired by unstructured Lumigraph rendering [Bue01], an *image-based warping method*, as well as a *refined geometric mapping technique for view-dependent projections onto geometrically and radiometrically complex surfaces* was presented [Bim05e]. It enables a *first projection of view-dependent content onto complex surfaces* (i.e., colored and textured surfaces with unknown geometry). Thereby, all geometric and radiometric compensation parameters are measured from a discrete number of multiple source camera positions during an offline calibration step. Since all parameters are stored in two-dimensional data structures, image-based interpolation methods can be applied. New parameters are estimated for a novel destination perspective during run-time using a weighting scheme that is normally applied for unstructured Lumigraph rendering in addition. Besides this image-based method, a *refined geometric mapping method* was presented. Both methods provide the required precision for radiometrically compensated view-dependent projections on complex surfaces, and support interactive frame rates (32-16 fps for displaying stereoscopic content with two projectors on an off-the-shelf hardware for a head-tracked observer). These methods enabled a *first stereoscopic visualization on ordinary surfaces within everyday environments*.

Camera-based geometric calibration techniques can be categorized into online or offline methods. While an offline calibration determines the calibration parameters (such as the projector-camera correspondence) in a separate step before runtime, an online calibration performs this task continuously during runtime. Much previous work has been carried out on offline calibration (including [Bim05e]) – but little on online techniques. Active offline calibration techniques usually rely on structured light projection to support enhanced feature detection [Pos82] [Cas98]. For simple surfaces with known geometry the geometric image

warping can be computed with beforehand determined constant calibration parameters. Examples are homography matrices (for planar surfaces [Yan01a, Che02]) or intrinsic and extrinsic projector parameters for non-trivial surfaces with known geometry [Ras99]. These techniques support a moving observer by warping the image in real-time with respect to the user's position. As explained above, for geometrically complex and textured surfaces with unknown geometry, projector-camera correspondences can be measured offline on a per-pixel basis for a discrete number of camera perspectives. During runtime, the correct warping is approximated in real-time by interpolating the measured samples depending on the observer's true perspective [Bim05e].

Online techniques can apply imperceptible structured light patterns that are seamlessly embedded into the projected image. This can be achieved by synchronizing a camera to a well selected time-slot during the modulation sequence of a DLP projector [Cot04]. Within this time-slot the calibration pattern is displayed and detected by the camera. Since such an approach requires modifying the original colors of the projected image, a loss in contrast can be an undesired side effect. Other techniques rely on a fast projection of images that cannot be perceived by the observer. This makes it possible to embed calibration patterns in one frame and compensate them with the following frame. Capturing altering projections of colored structured light patterns and their complements allows the simultaneous acquisition of the scene's depth and texture without loss of image quality [Was05].

A passive online method was described for supporting a continuous autocalibration on a nontrivial display surface [Yan01b]. Instead of benefiting from structured light projection, it directly evaluates the deformation of the image content when projected onto the surface. This approach assumes a calibrated camera-projector system and an initial rough estimate of the projection surface to refine the reconstructed surface geometry iteratively.

As an extension to the offline calibration method described in [Bim05e], a *hybrid (passive/active) calibration technique* was presented in [Zol06]. It corrects distortions that appear *online* when projecting images onto geometrically complex, colored and textured surfaces.

This method analyzes the optical flow that results from perspective distortions during motions of the observer and tries to use this information for computing the correct image warping on the fly. If this fails due to an unreliable optical flow analysis, an accurate –but slower and visible– structured light projection is automatically triggered. Together with an appropriate radiometric compensation, view-dependent content can be projected onto arbitrary everyday surfaces. An implementation mainly on the GPU ensures fast frame rates.

3.2.3 Defocus Compensation

Today's consumer projectors are designed and engineered to focus images on planar display surfaces. The Schleimpflug principle describes how to offset the focal plane by an off-axis configuration of the optical system. However, plane-focused images are partially blurred if projected onto surfaces with substantial depth differences. Special lenses, such as f-theta lenses, allow generating focused images on spherical surfaces. Planetariums and some cylindrical projection displays [Bie04] apply laser projectors to overcome this problem. Direct-writing-scanning-laser-beam projectors scan almost parallel beams of laser light onto the projection screen. Thereby, the laser beams remain constant in diameter over a substantial depth range. This results in a large focal depth and in the possibility to display sharp images in large dome-like or cylindrical theatres. The cost of a single laser projector, however, can quickly exceed the cost of several hundred conventional projectors. But the development of low-cost laser-diodes is promising and can overcome these drawbacks in future.

If the projection surface is multi-planar, multiple projectors can be arranged in such a way that they focus on individual planar sections (e.g. [Low01]). This, however, becomes inefficient and sometimes impossible the more complex the surface becomes.

A novel ***multi-focal projection*** technique [Bim06c] was developed that projects images with minimal defocus onto geometrically and radiometric complex surfaces. This is essential to enable, for instance, stereoscopic projections supporting disparity-based depth perception on arbitrary surfaces [Bim05e].

The multi-projector technique works as follows: Multiple conventional projectors with differently adjusted focal planes, but overlapping image areas are used. They can be either arbitrarily positioned in the environment, or can be integrated into a single projection unit. The defocus created on the surface is estimated automatically for each projector pixel via camera feedback and structured light projection. If this is known, a final image with minimal defocus can be composed in real-time from individual pixel contributions of all projectors.

The technique is independent of the surfaces' geometry, color and texture, the environment light as well as of the projectors' position, orientation, luminance and chrominance.

This method is the ***first rendering technique that compensates defocus effects caused by projectors***. Later, other methods have been presented that approach to enhance the overall sharpness of an image displayed by a single projector [Zha06, Bro06]. For these techniques, the defocus kernels of light samples projected onto complex scenes were analyzed. Based on these results, image sharpening was employed to compensate for optical defocus digitally. The limits of sharpening for such techniques, however, are set by the actual defocus of the projector and of the original image. Thus, it is not possible to compensate optical defocus if the original image does not contain a minimal amount of digital blur. This is not the case for the multi-projector method described in [Bim06c], since multiple physical focal planes are spanned by several projectors. Consequently, the focal depth increases with the number of projectors. The multi-focal projection technique in [Bim06c] is also ***the first (and currently only) defocus technique that supports multiple projectors simultaneously***. This method does require at least two projectors. A combination of single- [Zah06] and multi-projector [Bim06c] techniques would lead to an optimal tradeoff between image quality and system complexity.

3.2.4 Compensating Global Effects

Previously proposed image compensation techniques assume a simple geometric relation between cameras and projectors that can be automatically derived using structured light projections or co-axial projector-camera alignments. For projector-camera systems, this results in a precise mapping between camera and projector pixels. In reality, the light of a projected pixel often bounces back and forth several times at different areas on the surface, before it eventually reaches the imaging sensor of the camera. Due to interreflection, refraction, scattering and other global illumination effects, multiple camera pixels at spatially distant regions on the image plane may be affected by a single projector pixel. A direct mapping usually considers only camera pixels with the highest intensity contribution that result from the captured light of corresponding modulated projector pixels. Consequently, all global illumination effects are discarded. In some cases it might not even be possible to acquire a direct mapping at all because global effects are too dominant. Thus, all existing radiometric compensation techniques, for instance, consider local illumination effects (i.e., direct projections) only. Two solutions to this problem have been presented:

The ***compensation of surface-to-surface scattering*** for diffuse surfaces with known screen geometry using a ***reverse radiosity*** technique was presented in [Bim06a, Bim06b].

The radiometric compensation techniques described above can neutralize artifacts effects that result from a direct illumination only. However, diffuse surfaces scatter a fraction of light to other surface portions. This amount of indirect illumination adds to the direct illumination and has to be compensated as well. Problematic is the fact that the amount of scattering depends on the directly projected image and vice versa. A numerical and an analytical method have been presented that compensate the amount of scattered light in real-time.

The *numerical method* is implemented via pixel shading and multi-pass rendering which together realizes a Jacobi solver for sparse matrix linear equation systems [Bim06a].

It was recently shown by Seitz et al. [Sei05] that global interreflections can be removed from photographs of unknown scenes under unknown illuminations by applying an interreflection cancellation operator. It was shown in [Bim06b] that if the interreflection cancellation operator is applied to an image that does not contain interreflections, scattering can also be fully neutralized. This also proves that the amount of indirect scattering which is produced by the compensated image throughout all scatter levels equals the amount of indirect scattering that is produced by the original image in the first level. Since the interreflection cancellation operator is represented as a matrix, the *analytical compensation of scattering* in projected images can be realized as a simple matrix vector multiplication [Bim06b].

Both methods support real-time frame rates and can be applied together with *monoscopic and stereoscopic, immersive and semi-immersive projection screens*, such as CAVEs, workbenches, domes, and cylinders for displaying *more brilliant and uniform images with a reduced amount of scattered light*.

Acquiring the forward light transport between a projector and a camera allows capturing the entire light modulation of a projected pattern within an arbitrarily complex scene - including all local and global illumination effects [Sen05]. It represents a 4D slice of the 8D reflectance field and can be described as a linear equation system.

Replacing the camera image in the forward light transport equation with a desired picture enables to perform radiometric compensation by solving this equation system for the corresponding projector pattern [Wet06].

To achieve this in practice, the transport matrix is decomposed into a set of independent clusters mutually influencing projector and camera pixels that form smaller equation systems which can finally be processed. The compensation is modeled as a linear system that can be solved with respect to the projector patterns. Solving the set of equation systems explicitly on the CPU, however, does not support a compensation in real time. Pre-computing the inverse light transport matrix on the CPU and performing a simple matrix-vector multiplication during run-time on the GPU does lead to interactive frame rates.

Consequently, this novel method performs an *image based radiometric compensation which accounts for all possible types of light modulation*. It *unifies existing approaches that address individual problems* [Nay03, Bim05a, Par05, Sei05, Zha06, Bro06, Bim06a, Bim06b, Bim06c]. Based on examples, it was shown that it is possible to project corrected images onto complex surfaces such as an interreflecting statuette, glossy wallpaper, or through highly-refractive glass. Furthermore, it was illustrated that a side-effect of this approach is an increase in the overall sharpness of defocused projections. It can be expected, however, that other modulation types, such as diffractions, sub-surface scattering, or specular highlights will follow the same principle.

4 Future Perspectives

Future projectors will be small, portable, bright, extremely fast, high-resolution and high-contrast low-power units with integrated sensors. They will become components of mobile

devices, such as cell-phones and digital cameras, but will also find established places in our homes – as parts of game consoles and other home entertainment equipment. Appropriate display and illumination techniques –such as the ones described in this document– are as important as hardware improvements. Together, they will solve existing problems but will also enable new applications. The rapid developments of graphics hardware will support this evolution – and might even make its way into projector hardware.

Currently, the techniques that have been described in this document found *several applications outside the laboratories*. Some of them became inherent parts of *museum exhibitions, live stage performances, and multimedia presentations in historic sites*. At the moment, only those niche markets are addressed. With improving hardware and declining costs, however, mass applications might not be too far off.

Besides the continuous investigation of new techniques as well as the improvements of existing methods, one of the main challenges will be to *up-scale the projector-based illumination and display concepts* that have been described. The two initiatives that are outlined below will found the basis for near future research activities.

4.1 Spatial Augmented Reality for Architecture

Immersive and semi-immersive projection displays, such as CAVEs, walls, workbenches, cylinders, and domes are being used to support virtual reality applications in many professional domains. The visualization of data with these displays, however, requires dedicated rooms for setting up non-mobile screens, and allows the interaction with purely synthetic information only.

The aim of these research activities is to investigate and develop the conceptual and technological fundamentals for realizing such visualizations in real world environments. They strive for *enabling immersive and semi-immersive virtual reality, as well as augmented reality experiences without the need for special display surfaces or permanent screen configurations*. As a result the main *focus lies on the visualization of, and the interaction within existing buildings in the course of building surveying and planning processes* – investigating novel ways of on-site surveying, visualization and simulation.

Based on the evolving core technology, new ways for interactive information visualization are researched – supporting early architectural design phases. The vision is an ad hoc visualization of interactive stereoscopic three-dimensional and monoscopic two-dimensional data, such as lighting simulations, on arbitrary surfaces within real-world indoor environments. The primary question is: Which new possibilities and application areas are opened through this approach for the surveying and planning process within existing buildings? Particular attention should be given to user interfaces and interaction possibilities, as well as to an appropriate integration within the working process of architects.

This project is supported by the Deutsche Forschungsgemeinschaft (DFG) under contract number PE 1183/1-1.

4.2 Digital Illumination for Augmented Studios

Many modern TV productions apply virtual studio technology. A good overview can be found at [Gib98]. Chromakeying is the principle method for superimposing the live captured or recorded video signal of a physical blue screen studio with virtual content. Thereby, the video signal is analyzed and video pixels with a predefined color (e.g. blue or green) are replaced by computer-generated graphics. This allows using image processing techniques to efficiently

separate the foreground from the background, and consequently to integrate real objects (such as an actor or a moderator) seamlessly into a purely virtual environment.

Blue screen techniques, however, limit virtual studio technology to special recording environments. Therefore, recent research initiatives investigate the potential of augmented reality for TV productions. In contrast to blue screen studios, fully equipped real television studios are augmented with virtual content by superimposing the recorded video stream with computer graphics. According to virtual studios, this can be referred to as “augmented studios”.

Several groups have already shown the advantages of augmented reality in a studio production context: Yama et al. [Yam02], for instance, augment 360° ultra high-definition omni-directional images of artificial backgrounds –being distorted in real-time relative to the rotation of a pan-tilt camera, and being occluded by a real actor. An axi-vision camera [Kaw00] is used for simultaneously capturing color and depth information per pixel.

Recent examples are also shown in the context of the EU funded project MATRIS [Mat04]: Frahm et al. [Fra05] use a fish eye camera in addition to a studio camera. While the studio camera records the video content to be augmented, the fish eye camera observes the upper hemisphere to track the installed studio lights. Applying a structure from motion algorithm [Koc05] to both images makes the estimation of the studio camera’s pose possible. Standard stereo algorithms [Koc98] allow reconstructing the depth of the studio setting and consequently enable correct occlusion effects between real and virtual objects. Furthermore, the knowledge of the real studio light sources allows computing a light map that ensures a consistent illumination and shadowing.

Virtual and augmented studio productions have to solve several technical challenges. One of them is the robust and fast tracking of the studio cameras [Ber03]. Some approaches apply special tracking hardware, while others try to estimate the cameras’ pose by observing natural features (e.g., ceiling-mounted studio lights or the studio content itself) or artificial tags [Tho97] with additional cameras, as explained above. Optical tracking is becoming more and more popular due to its robustness against most environmental disturbances, speed and precision. Optical tracking approaches can be categorized into marker-less and marker-based methods. Marker-less techniques, on the one hand, rely strongly on the robust detection of natural scene features [Fra05]. They will fail for uniformly structured surfaces or under dim lighting conditions. This limits the application of such techniques in TV studios to optimized situations. Marker-based tracking, on the other hand, provides artificial visual features by integrating detectable marker tags. However, these markers should neither be directly visible within the studio environment nor appear in the recorded video stream. Consequently, marker-based tracking is usually restricted to observing out-shot areas –such as the ceiling or the floor– which are normally covered by studio equipment, like light installations, cables, and mountings. Thus, occlusions and dynamic reconfigurations of the installations cause additional problems for marker-based tracking.

Another problem is the acquisition of the scene depth. This is necessary for integrating synthetic 3D objects into the video stream while producing consistent occlusion and illumination effects with the recorded real content. Some approaches reconstruct the scene geometry offline (during a special calibration step) using multi-viewpoint stereo from uncalibrated video sequences. The quality of such techniques relies on the quality of feature matching in the stereo pairs. However, finding matchable features to support a high quality depth reconstruction might be difficult – not only for real studio environments, but also for virtual studios or embedded blue screens that mainly apply uniformly colored matting surfaces. Besides offline reconstruction of the static studio setting, online depth estimations (e.g., of moving people in the scene) is even more problematic.

Yet another challenge for virtual and augmented studios is the question of how to display direction information to moderators, actors or participants during a live broadcast or a

recording. Teleprompters or fixed screens offer limited possibilities since they do not allow bringing the presented information into a spatial context. Step sequences, for instance, are usually marked statically on the floor ground.

The application of *digital light projection for studio illumination* – both exclusively or in combination with an existing analog lighting – has been proposed in [Bim06f], and early proof-of-concept examples have been provided. This concept can solve several of the problems that are mentioned above, but also opens new possibilities for modern television productions:

- *dynamic re-illumination* of studio settings, moderators and actors without physical modification of the lighting equipment;
- *marker-based in-shot tracking* of studio cameras without visible markers;
- *dynamic presentation* of un-recorded direction, moderation and other information *spatially anywhere* within the studio;
- *integration of imperceptible coded patterns* that support continuous online-calibration, camera tracking, and acquisition of scene depth.

This project is submitted to the Deutsche Forschungsgemeinschaft (DFG) under reference number BI 835/2-1.

Acknowledgments

Special thanks to all students at the Bauhaus-University Weimar who contributed, or are still contributing to this work: Andreas Emmerling, Gordon Wetzstein, Anselm Grundhöfer, Christian Nitschke, Daniel Danch, Petro Kapakos, Thomas Klemmer, Franz Coriand, Alexander Kleppe, Stefanie Zollmann, Tobias Langlotz, Erich Bruns, Uwe Hahne, Mathias Möhring, Thomas Zeidler, Sebastian Knödel, Ferry Häntsch, Daniel Kolster, Benjamin Schmidt, Emanuel Züger, Saskia Groenewegen, Tim Gollub, and Daniel Kurz.

References

The highlighted publications are attached to this document. Note that the work presented in [Bim06d] does not count as a publication. It has been attached only as a complement to [Bim06e].

[And89] Andrews, J.R. and Haas, W.H., Compact Illuminators for Transmission Holograms, Proceedings of SPIE' 89, Practical Holography III, vol. 1051, Springer, pp. 156-159, 1989

[Ash06] Ashdown, M., Okabe, T., Sato, I. and Sato, Y., Robust Content-Dependent Photometric Projector Compensation, Proc. IEEE International Workshop on Projector-Camera Systems, pp. 60-68, 2006

[Ber03] Bernas, M., Teisler, M. Determination of the camera position in virtual studio Signal Processing and Information Technology. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on 14-17 Dec. pp. 535 – 538, 2003

[Bie04] Biehling, W., Deter, C., Dube, S., Hill, B., Helling, S., Isakovic, K., Klose, S., and Schiewe, K, LaserCave – Some Building Blocks for immersive Screens. Proc. of Int. Status Conference on Virtual - and Augmented Reality, 2004

[Bim00] Bimber, O., Encarnação, L.M., and Schmalstieg, D. Augmented Reality with Back-Projection Systems using Transflective Surfaces. Computer Graphics Forum (proceedings of Eurographics 2000 - EG'2000), vol. 19, no. 3, pp. 161-168, 2000

[Bim01a] Bimber, O., Encarnação, L.M. and Branco, P. The Extended Virtual Table: An Optical Extension for Table-Like Projection Systems. Presence: Teleoperators and Virtual Environments, vol.10, no. 6, pp. 613-631, 2001

[Bim01b] Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnação, L.M. The Virtual Showcase. IEEE Computer Graphics and Applications, vol. 21, no.6, pp. 48-55, 2001

[Bim02] Bimber, O. and Fröhlich, B. Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays. In proceedings of International Symposium on Mixed and Augmented Reality (ISMAR'02), pp. 186-195, 2002

[Bim03a] Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnação, L.M. Real-Time View-Dependent Image Warping to correct Non-Linear Distortion for Curved Virtual Showcase Displays. Computers and Graphics - The international Journal of Systems and Applications in Computer Graphics, vol. 27, no.4, pp. 512-528, 2003

[Bim03b] Bimber, O., Grundhöfer, A., Wetzstein, G., and Knödel S. Consistent Illumination within Optical See-Through Augmented Environments. In proceedings of IEEE/ACM International Symposium on Augmented and Mixed Reality (ISMAR'03), pp. 198-207, 2003

[Bim04a] Bimber, O. Combining Optical Holograms with Interactive Computer Graphics. In IEEE Computer, pp. 85-91, January 2004

[Bim04b] Bimber, O. HoloGraphics: Combining holograms with interactive computer graphics. *The Holographer*, 2004

[Bim05a] Bimber, O., Coriand, F., Kleppe, A., Bruns, E., Zollmann, S., and Langlotz, T. Superimposing Pictorial Artwork with Projected Imagery. *IEEE MultiMedia*, pp. 16-26, January-March issue 2005

[Bim05b] Bimber, O., Emmerling, A., and Klemmer, T. Embedded Entertainment with Smart Projectors. *IEEE Computer*, pp. 56-63, January issue 2005

[Bim05c] Bimber, O., Combining Optical Holograms with Interactive Computer Graphics, Holography, SPIE International Technical Group Newsletter, vol. 16, no. 1, pp. 1+9, June 2005

[Bim05d] Bimber, O., Zeidler, T., Grundhöfer, A., Wetzstein, G., Möhring, M., Knödel, S., and Hahne, U. Interacting with Augmented Holograms. In proceedings of SPIE Conference on Practical Holography XIX: Materials and Applications, 2005

[Bim05e] Bimber, O., Wetzstein, G., Emmerling, A., and Nitschke, C. Enabling View-Dependent Stereoscopic Projection in Real Environments. In proceedings of International Symposium on Mixed and Augmented Reality (ISMAR'05), pp. 14-23, 2005

[Bim06a] Bimber, O., Grundhöfer, A., Zeidler, T., Danch, D., and Kapakos, P. Compensating Indirect Scattering for Immersive and Semi-Immersive Projection Displays. In proceedings of IEEE Virtual Reality (IEEE VR'06), pp. 151-158, 2006

[Bim06b] Bimber, O., Emerging Technologies of Augmented Reality: Interfaces & Design, Chapter on Projector-Based Augmentation, Idea Group (publisher), Haller/Thomas/Billinghurst (editors). To appear in 2006

[Bim06c] Bimber, O. and Emmerling, A., Multi-Focal Projection: A Multi-Projector Technique for Increasing Focal Depth. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 12, no. 4, pp. 658-667, 2006

[Bim06d] Bimber, O., Augmenting Holograms. To appear in *IEEE Computer Graphics and Applications (CG&A)*, September 2006

[Bim06e] Bimber, O., Merging Graphics and Holograms. To appear in *Journal of Holography and Speckle*, October 2006

[Bim06f] Bimber, O., Grundhöfer, A., Zollmann, S., and Kolster, D., Digital Illumination for Augmented Studios. To appear in the *Journal of Virtual Reality and Broadcasting*, October 2006

[Bro06] Brown, M., Song, P., and Cham, T.-J., Image Pre-Conditioning for Out-of Focus Projector Blur. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006

[Bue01] Buehler, C., Bosse, M., McMillan, L., Gortler, S.J., and Cohen, M.F., Unstructured Lumigraph Rendering, *Proc. of ACM Siggraph'01*, pp. 425-432, 2001

- [Cas98] Caspi, D., Kiryati, N., Shamir, J.: Range imaging with adaptive color structured light, IEEE Transactions on Pattern analysis and machine intelligence, pp. 470-480, Volume 20, 1998
- [Che02] Chen, H., Sukthankar, R., and Wallace, G., Scalable alignment of large-format multi projector displays using camera homography trees, Proc. of IEEE Visualization, pp. 339-346, 2002.
- [Cot04] Cotting, D., Naef, M., Gross, M., Fuchs, H., Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display, IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR04), pp. 100-109, Arlington, 2004
- [Dre97] Dreesen, F., and von Bally, G., Color Holography in a Single Layer for Documentation and Analysis of Cultural Heritage, In Proceedings of Optics within Life Sciences (OWLS IV), Springer, pp. 79-82, 1997
- [Dre00] Dreesen, F., Deleré, H., and von Bally, G., High-Resolution Color Holography for Archaeological and Medical Applications, In Proceedings of Optics within Life Sciences (OWLS V), Springer, pp. 349-352, 2000
- [Ehn04] Ehnes, J., Hirota, K., and Hirose, M., Projected Augmentation – Augmented Reality using Rotatable Video Projectors, Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 26-35, 2004
- [Fra05] Jan-Michael Frahm, Kevin Koeser, Daniel Grest, Reinhard Koch: Markerless Augmented Reality with Light Source Estimation for Direct Illumination. CVMP, London, December 2005
- [Fuj05] Fujii, K., Grossberg, M.D., Nayar, S.K., A Projector-Camera System with Real-Time Photometric Adaptation for Dynamic Environments, IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'05, Volume 01, pp. 814-821, 2005
- [Gib98] Gibbs, S., Arapis, C., Breiteneder, C. Lalioti, V. Mostafawy, S. Speier, J. Virtual studios: an overview. IEEE Multimedia. vol.5, no.1, pp. 18-35, 1998
- [Gros04] Grossberg, M.D., Peri, H., Nayar, S.K., and Bulhumeur, P., Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 1, pp. 452-459, 2004
- [Gru06] Grundhöfer, A. and Bimber, O., Real-Time Adaptive Radiometric Compensation. Submitted to Transactions on Visualization and Computer Graphics (TVCG), 2006**
- [Kaw00] Kawakita, M. et al. 2000. Axi-vision camera: a three-dimension camera, In Three-Dimensional Image Capture and Applications III. Proceedings of SPIE, Vol.3958, pp. 61-70, Jan. 2000

- [Kiy00] Kiyokawa, K., Kurata, Y. and Ohno, H. An Optical Seethrough Display for Mutual Occlusion of Real and Virtual Environments. In proceedings of IEEE & ACM ISAR 2000, pp. 60-67, 2000
- [Koc98] R. Koch, M. Pollefeys, and L. Van Gool. Multi viewpoint stereo from uncalibrated video sequences. In Proceedings of ECCV'98, vol. 1 of LNCS, pp. 55-71, Freiburg, Springer-Verlag, 1998
- [Koc05] Markerless Image-based 3D Tracking for Real-time Augmented Reality Applications. R.Koch, K.Koeser, B.Streckel, J.-F. Evers-Senne WIAMIS 2005, April 13-15, Montreux, Switzerland, 2005
- [Kol89] Kollin, J.S., Benton, S.A., and Jepsen, M.L., Real-Time Display of 3-D Computed Holograms by Scanning the Image of an Acousto-Optic Modulator, Proceedings of SPIE, vol. 1136, Holographic Optics II: Principles and Applications, Springer, pp. 178-185, 1989
- [Low01] Low, K-L., Welch, G., Lastra, A., and Fuchs, H., Life-Sized Projector-Based Dioramas, Proc. Symp. Virtual Reality Software and Technology (VRST'01), 93-101, 2001
- [Luc93] Lucente, M., Interactive Computation of Holograms Using a Look-Up Table, Journal of Electronic Imaging, vol. 2, no. 1, pp. 28-34, 1993
- [Luc95] Lucente, M. and Tinsley, A., Rendering Interactive Images, Proceedings of ACM Siggraph'95, ACM Press, pp. 387-394, 1995
- [Luc97] Lucente, M., Interactive Three-Dimensional Holographic Displays: Seeing the Future in Depth, ACM Computer Graphics, Special Issue on Current, New, and Emerging Display Systems, ACM Press, pp. 63-67, 1997
- [Mat04] Markerless real-time Tracking for Augmented Reality Image Synthesis, <http://www.ist-matris.org/>, 2004
- [Nae02] Naemura, T., Nitta, T., Mimura, A., Harashima, H. Virtual Shadows – Enhanced Interaction in Mixed Reality Environments. In proceedings of IEEE Virtual Reality (IEEE VR'02), pp. 293-294, 2002
- [Nay03] Nayar, S.K., Peri, H., Grossberg, M.D., and Belhumeur, P.N., A Projection System with Radiometric Compensation for Screen Imperfections, Proc. of International Workshop on Projector-Camera Systems, 2003
- [Nod99] Noda, S., Ban, Y., Sato, K., and Chihara, K. An Optical See-Through Mixed Reality Display with Realtime Rangefinder and an Active Pattern Light Source. Transactions of the Virtual Reality Society of Japan, vol. 4, no. 4, pp. 665-670, 1999
- [Oka97] Okamoto, M., Ueda, H., Nakamura, I., Shimizu, E., and Kubota, T., Multiple Illuminations Method for Rainbow Holograms using an LCD Projector, Proceedings of SPIE'97, Practical Holography XI, vol. 3011, Springer, pp. 70-75, 1997
- [Par06] Park, H., Lee, M.-H., Kim, S.-J., and Park, J.-II, Contrast Enhancement in Direct Projected Augmented Reality, ICME'06: International conference on Multimedia and Expo, 2006

- [Pet03] Petz, C. and Magnor, M., Fast Hologram Synthesis for 3D Geometry Models Using Graphics Hardware, Proceedings of SPIE'03, Practical Holography XVII and Holographic Materials IX, vol. 5005, Springer, pp. 266-275, 2003
- [Pin01] Pinhanez, C. The everywhere displays projector: A device to create ubiquitous graphical interfaces, In proceedings of Ubiquitous Computing, pp. 315-331, 2001
- [Pos82] Posdamer, J. L., Altschuler, M. D., Surface measurement by space-encoded projected beam systems, Computer Graphics and Image Processing, pp. 1-17, vol. 18 issue 1, 1982
- [Ras99] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., and Fuchs, H., Multi-projector displays using camera-based registration, Proc. of IEEE Visualization, pp. 161-168, 1999
- [Ras01] Raskar, R. Welch, G., Low, K.L., and Bandyopadhyay, D. Shader Lamps: Animating real objects with image-based illumination. In Proc. of Eurographics Rendering Workshop (EGRW'01), 2001
- [Ras03] Raskar, R., van Baar, J., Beardsly, P., Willwacher, T., Rao, S. and Forlines, C. iLamps: Geometrically Aware and Self-Configuring Projectors, In proceedings of ACM Siggraph, pp. 809-818, 2003
- [Sei05] Seitz, S.M., Matsushita, Y., & Kutulakos, K.N. (2005). A Theory of Inverse Light Transport. Proc. of IEEE International Conference on Computer Vision (IEEE ICCV'05), 1440-1447, 2005
- [Sen05] Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., and Lensch, H. P. A., Dual Photography. ACM Trans. Graph. 24, 3, 745–755, 2005
- [Sli04] Slinger, C.W., Cameron, C.D., Coomber, S.J., Miller, R.J., Payne, D.A., Smith, A.P., Smith, M. A. G., and Stanley, M., Recent developments in computer generated holography: towards a practical electroholography system for interactive 3D visualization, Proceedings of SPIE'04, Practical Holography XVIII, vol. 5290 ,pp. 27-41, 2004
- [Sli05] Slinger, C. W., Cameron, C. D., and Stanley, M., Computer-generated holography as a generic display technology, IEEE Computer, vol 38, no. 8, pp. 46-53, 2005
- [Tho97] Thomas, G.A. Jin, J. Urquhart, C. A versatile camera position measurement system for virtual reality tv production. International Broadcasting convention. Amsterdam. 12-16 September, pp. 284-289, 1997
- [Und99] Underkoffler, J., Ullmer, B., and Ishii, H., Emancipated Pixels: Real-World Graphics in the Luminous Room, In proceedings of SIGGRAPH 99, Computer Graphics Proceedings, Annual Conference Series, edited by Alyn Rockwood, pp. 385–392. Reading, MA: Addison Wesley Longman, 1999
- [Vil93] Vila, D., Folding Time into Space – Novel Approaches to Integral Holography, Proceedings of SPIE'93, Practical Holography VII: Imaging and Materials, vol. 1914, Springer, pp. 230-235, 1993

[Vil94] Vila, D., Holo-Dynamics – Linking Holography to Interactive Media, Proceedings of SPIE'93, Practical Holography VIII, vol. 2176, Springer, pp. 166-171, 1994

[Wan05] Wang, D., Sato, I., Takahiro, O. and Sato, Y., Radiometric Compensation in a Projector-Camera System Based Properties of Human Vision System, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, Volume 03, page 100, 2005, 2005

[Was05] Waschbüsch, M., Würmlin, S., Cotting, D., Sadlo, F., Gross, M., Scalable 3D Video of Dynamic Scenes. The Visual Computer, Volume 21, pp. 629-638, 2005

[Wet06] Wetzstein, G. and Bimber, O., A Generalized Approach to Radiometric Compensation. Submitted to ACM Transactions on Graphics, 2006

[Yam98] Yamasakia, K., Okamotoa, M., Nakamura, I., and Shimizu, E., Fluctuation Hologram with Multiple Images, Proceedings of SPIE'98, Practical Holography XII, vol. 3293, Springer, pp. 139-144, 1998

[Yam02] Yamanouchi, Y. et al. Real space-based virtual studio seamless synthesis of a real set image with a virtual set image. Proceedings of the ACM symposium on Virtual reality software and technology. Hong Kong, China. SESSION: Hybrid VR. pp. 194-200, 2002

[Yan01a] Yang, R., Gotz, D., Hensley, J., Towles, H., PixelFlex: A Reconfigurable Multi-Projector Display System. IEEE Visualizations, pp. 161-168, 2001

[Yan01b] Yang, R., Welch, G., Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery. Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision, Plzen, 2001

[Yos03] Yoshida, T., Horii, C., and Sato, K., A Virtual Color Reconstruction System for Real Heritage with Light Projection, Proc. Virtual Systems and Multimedia, IEEE CS Press, pp. 158-164, 2003

[Zha06] Zhang, L. and Nayar, S. K., Projection Defocus Analysis for Scene Capture and Image Display. ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH), 2006

[Zol06] Zollmann, S., Langlotz, T. and Bimber, O., Passive-Active Geometric Calibration for View-Dependent Projections onto Arbitrary Surfaces, To appear at Workshop on Virtual and Augmented Reality of the GI-Fachgruppe AR/VR, September 2006

Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays

Oliver Bimber[#]

Fraunhofer Center for Research in Computer Graphics, 321 South Main St., Providence, RI 02903, USA,
oliver.bimber@medien.uni-weimar.de

[#]now at Bauhaus University Weimar

Bernd Fröhlich

Bauhaus University Weimar
Bauhausstraße 11,
99423 Weimar, Germany,
bernd.froehlich@medien.uni-weimar.de

Abstract

This paper presents projector-based illumination techniques for creating correct occlusion effects for optical see-through setups. We project view-dependent occlusion shadows onto the real surfaces that are located behind virtual objects. This results in a perfect occlusion of real objects by virtual ones. We have implemented and tested our approach in the context of the Virtual Showcase display. We describe hardware extension for projecting light into the showcase and present our rendering techniques for displaying occlusion shadows for single and multi-user environments as well as for single and multi-light-projector configurations. We also report on the limitations of our system for multi-user situations and describe our experiences with a first experimental prototype.

1. Introduction

Projection-based augmented reality systems, such as the Virtual Showcase [3], share many positive properties of projection-based virtual environments. These displays provide high resolution, improved consistency of eye accommodation and convergence, little motion sickness potential, and the possibility of an integration into common working environments. One of the main challenges for projection-based AR systems as well as for head-mounted optical see-through displays is the generation of correct occlusion effects between virtual and real objects [1]. Additionally shadows of virtual objects cast onto real ones and consistent illumination of the real and virtual scenery are often difficult to achieve.

In this paper, we introduce projector-based illumination techniques for view-dependent optical see-through AR displays. This approach has the potential to solve all of the above mentioned problems. Here, we focus on using projector-based illumination for creating

correct occlusion effects for mixed reality configurations (cf. figure 1).

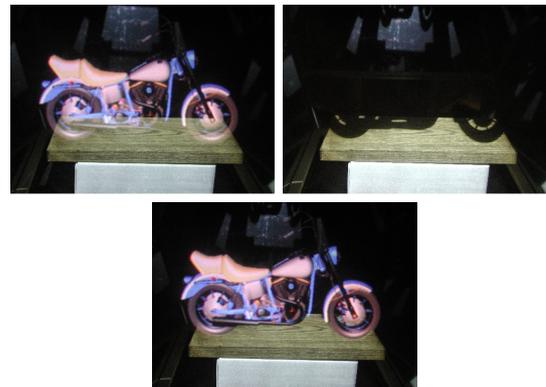


Figure 1: Wrong occlusion effects with normal illumination (left), occlusion shadow generated with projector-based illumination (right), realistic occlusion of the real object by the virtual one (center).

We have implemented and tested such a system in the context of the Virtual Showcase, which consists of a horizontal projection screen and a convex half-silvered mirror assembly (cf. figure 2). Virtual and real objects can be displayed in the same space inside the showcase.

The original Virtual Showcase used a standard light bulb to illuminate real objects. This setup does not provide very much control over the lighting situation. By using a computer-controlled video-projector as a replacement for the simple light bulb, we are able to fully control the lighting situation inside the showcase on a per-pixel basis.

Our main contribution is a solution to the problem of correct occlusion for mixed reality scenarios with view-dependent optical see-through displays. Our method produces correct occlusion effects between virtual and

real objects by projecting shadows onto real objects located behind virtual ones using projector-based illumination.



Figure 2: Our experimental prototype: The Virtual Showcase sits on top of a rear-projection screen. The lower half of the truncated pyramid configuration consists of four half-silvered mirrors. An additional set of full mirrors comprise the top half and redirects the light beam of a video projector (upper right) into the showcase center.

We describe our extended Virtual Showcase hardware for projecting these shadows into the showcase (cf. figure 2) and present rendering techniques for displaying them in single and multi-user environments as well as for single and multi-light-projector configurations. We also report on the limitations and potential extensions of our system and describe our experiences with our first setup.

2. Related Work

Kiyokawa et al. [8] present ELMO, an optical see-through head-mounted display that supports mutual occlusion. ELMO uses half-silvered mirrors as optical combiners and an additional semi-transparent LCD panel in front of the conventional optics. The LCD panel is used to selectively block the incoming light on a per-pixel basis. This enables virtual objects to occlude real ones. A head-attached depth sensor allows them to acquire depth maps of the real environment in real time. This makes the occlusion of virtual objects by real ones possible. ELMO faces a number of problems that are linked to the LCD panel: light attenuation caused by the LCD panel, and low response time and resolution of the LCD panel. However, as the first functioning system of its kind, it effectively addresses the occlusion problem of optical see-through head-mounted displays.

Noda et al. [10] present a stationary optical see-through display that uses a video projector to illuminate real objects selectively – not lighting those areas that are overlaid by graphics. Noda’s system is strictly limited in

several points. Firstly a dark surrounding environment is required, which constrains the applications possible. Secondly view-dependent rendering is not possible. The observer’s viewpoint has to match with the center of projection of the video projector since the illumination pattern is rendered from this point using a normal on-axis projection. In this special case no depth information of the real environment is required for a correct rendering. Lastly stereoscopic rendering is not provided.

Naemura et al. [9] proposes an approach that is technically similar to Noda’s. The conceptual difference, however, is that he applies a hand-held video projector as a real flashlight to interactively generate shadow effects of virtual objects on real surfaces. He does not address the occlusion problem of optical see-through displays, but focuses on enhancing such interactive mixed reality applications by providing additional visual cues through shadows. As in Noda’s case no depth information of the real objects are needed.

Head-Mounted Projective Displays, or HMPDs, (such as described by Hua et al. [7]) require the observer to wear miniature projectors. The projectors beam the synthetic images directly onto the surfaces of the real objects that are within the user’s field of view. Since the observer’s viewing frustum can be optically matched with the projection frustum, view-dependent rendering is possible while benefiting from a view-independent projection (i.e., depth information for real objects is not required). However, the real objects’ surfaces have to be coated with a retro-reflective material in terms of providing stereoscopic rendering, multi-user applications, and the usage of such displays within uncontrolled illuminated environments. The occlusion problem of optical see-through displays is not an issue for HMPDs, since the retro-reflective material avoids the problem of environment light interfering with the graphical overlays.

Raskar et al. [12] applies multiple stationary video projectors to “lift” the lighting and material properties of real objects by projecting colored images onto the real objects’ surfaces. His approach provides an auto-stereoscopic behavior and does not have to deal with the occlusion problem since it is not based on the optical see-through concept. In fact, he faces an inverse problem: His method is constrained by the shape and color of the real objects. On the one hand, it is not possible to display graphics next to a real surface if another real object is not located behind the graphics that can serve as display surface. On the other hand, the real objects are required to have a bright color that diffuses the projected light. Dark objects would absorb the light. However, since a view-dependent rendering is mostly not required¹, he can

¹ Basic view-dependent illumination effects, such as specular reflection, are handled by a skillful distribution of tasks between model-view and projection transformations.

simply render a textured virtual representation of the real scene from the viewpoint of the projector(s).

3. Our Approach

Being an optical see-through display, the Virtual Showcase faces the same occlusion problem as head-mounted displays if conventional illumination is used. However, the Virtual Showcase completely encloses the contained real artifact, which offers the possibility to fully control the lighting situation. Figure 2 shows our experimental setup with an additional video-projector for illuminating the real content inside the showcase. We refer to these projectors as *light projectors*.

Our idea is similar to Noda’s [10]. However, our approach supports view-dependent and stereoscopic rendering for single and multiple users and we do not require the illumination being projected from the user’s point of view. This requires depth knowledge of the real scenery to support both –the occlusion of real objects by virtual ones and vice versa. In addition, the Virtual Showcase setup does not depend on a dark surrounding, since the real artifact is completely enclosed and the interior lighting of the Virtual Showcase is fully controllable.

We dynamically generate shadows directly on the real objects’ surfaces wherever graphics is overlaid (figure 1). These shadows are not directly visible to the observers, since they are purposely occluded by the overlaid graphics. We call these shadows *occlusion shadows*. We additionally render phantom bodies representing real objects which occlude virtual objects behind them. The combination of occlusion shadows and phantom bodies effectively solves the occlusion problem for optical see-through displays such as the Virtual Showcase.

4. Rendering Occlusion Shadows

For rendering occlusion shadows the viewpoints of each user, the intrinsic and extrinsic parameters of each light projector, as well as the virtual and the real scene must be known.

The viewpoints are continuously measured with head-tracking technology, while the light projectors’ parameters are determined only once during a calibration phase. Virtual objects can be interactively manipulated within the showcase during runtime.

Knowing the scene and the view transformation lets us compute the perspective projection matrix (V) of the corresponding viewpoint that incorporates the model-view transformation with respect to the scene’s origin

4.1. Light Projector Calibration

Before calibrating a light projector, a geometric representation of the real scene is registered to its physical counterpart. Then, the two-dimensional perspective projections of selected three-dimensional points on the real objects’ surfaces are sampled within the light projector’s screen space as described by Raskar [12]. The three-dimensional fiducials are highlighted on the real objects’ surfaces by rendering and overlaying them with the Virtual Showcase display – given that the real objects have been registered first. A crosshair is then rendered into the light projector’s frame buffer. It is aligned with the highlighted surface points to measure their 2D projections in the corresponding screen space.

Users interactively browse through the sample points, which allows the selection of reasonable calibration areas (e.g., those that are clearly visible and are not in the shadow of, or occluded by other surfaces). Once an appropriate number of samples has been taken, they are used as input for a numerical minimization which computes the light projector’s intrinsic (vertical field of view and aspect ratio in our case) and extrinsic (position, optical axis, and up-vector in our case) parameters. We applied Powell’s direction set method [11] to solve this perspective-n-point (PnP) problem. The result is the projector’s perspective projection matrix (P) that incorporates the correct model-view transformation with respect to the scene origin.

In our case the light frustum of the projector is redirected by a planar mirror. Thus we need to incorporate the reflection transformation of the mirror. During calibration, we reflect the coordinates of the 3D fiducials over the corresponding mirror plane before passing them into the minimization routine. In this case, P needs to incorporate an additional reflection matrix that reflects the scene over the mirror plane during rendering (as described in Bimber et al. [4]).

If multiple projectors are used, the calibration process has to be repeated for each projector separately.

4.2. Single Viewpoint

The basic algorithm below illustrates how to render occlusion shadows for a single point of view.

The depth information of both – the real and the virtual content have to be known. A shadow mask that contains the silhouette of the virtual content is generated (lines 1-5) which is then perspectively mapped onto the known geometry of the real content (lines 6-7). Line 4 renders the illumination for the real content into the frame buffer. This illumination could be computed with a similar BRDF model as described in Raskar et al. [12] – producing a correct and matching radiance on real and virtual surfaces with respect to virtual light sources. Note

that this has not been implemented yet. We just project uniformly colored light onto the real surfaces from the light projector's point of view while virtual objects are illuminated from the positions of the virtual light sources.

- | | |
|----|--|
| 1: | set projection matrix to V |
| 2: | render real content into depth buffer |
| 3: | render virtual content into stencil buffer |
| 4: | render illumination for real content into frame buffer (previously cleared to black) |
| 5: | transfer frame buffer into texture memory T |
| 6: | set projection matrix to P , set texture matrix to V + normalization space correction, clear frame buffer to black |
| 7: | render real content into frame buffer using projective texture T |

Algorithm 1

Note also that the instruction in line 2 ensures a correct occlusion of virtual objects by real ones, as proposed by Breen et al. [5]. This is illustrated in figure 3. The normalization space correction in line 6 consists of a scaling by $[0.5, 0.5, 1.0]$, followed by a translation of $[0.5, 0.5, 0.5]$ to map from normalized screen space to normalized texture space².

4.3. Multiple Viewpoints

A clear limitation of our method is the following fact: If the same real surfaces are simultaneously visible from multiple points of view (e.g. for different observers), individual occlusion shadows that project onto these surfaces are also visible from different viewpoints at the same time.

Considering two observers, for instance, observer A might be able to see the occlusion shadow that is generated for observer B and vice versa. In addition, the shadows move if the viewers are moving, which might be confusing. This problem cannot be solved in general with our current setup. However, we propose two approaches to reduce these effects:

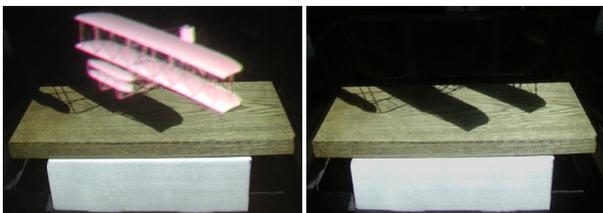


Figure 4: Occlusion shadows generated for two different viewpoints. With graphical overlay (left), and without graphical overlay (right).

² This applies for OpenGL.

Occlusion shadows generated for other viewpoints are the umbral hard-shadows that are cast by the virtual scene with a light source positioned at the other viewpoints' locations. We make use of this fact by attaching a point light to each viewpoint. This generates correct lighting effects on the virtual scene's surfaces – in addition to matching hard-shadows on the real scene's surfaces (cf. figure 4).

Our second approach tries to minimize the interference between individual occlusion shadows by ensuring that they are generated only on those real surfaces that are visible from the corresponding viewpoint. However, since the occlusion shadows are finally rendered from the viewpoint of the projector, all view-dependent computations (e.g., back-face culling and depth buffering) are done for this perspective – not for the perspectives of the actual viewpoints.

Figure 5 illustrates a simple example. Here, we assume two viewpoints ($V1$, $V2$), one light projector (P), and five real surfaces (1-5). For $V1$, surface 1,2 and 5 are completely visible, surface 3 is completely invisible, and surface 4 is partially visible. For $V2$, surfaces 2,3,4 and 5 are completely visible and surface 1 is completely invisible. Consequently, surfaces 2, 4 and 5 are (at least partially) visible for both viewpoints.

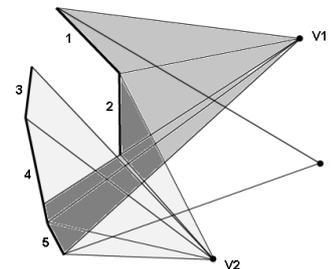


Figure 5: Visibility for different points of view.

We want to ensure that the occlusion shadows for each viewpoint are generated only on the visible portions of the surfaces. However, since different surfaces are visible from the perspective of the projector than from the perspectives of the viewpoints, the correct appearance of real scene has to be determined before it is rendered.

If algorithm 1 would be simply repeated for every viewpoint³, the projective texture of $V1$ would be unnecessarily mapped onto surface 4⁴ and might consequently interfere with the texture of $V2$. Algorithm 2 explains how we approach this problem.

Throughout lines 1-12 in algorithm 2, a shadow mask is generated for each viewpoint and stored in an

³ And the resulting textures would be color blended appropriately.

⁴ Note that surface 3 cannot be illuminated, since it isn't visible from the perspective of the projector.

individual block of the texture memory. All shadow masks are then color blended into the final image that is projected by the light projector (lines 14-20). To support a proper color blending, the first shadow map is rendered with a black shadow color and the assigned light color as background. It is used to create a base image during the first rendering iteration. All subsequent iterations generate masks with a black shadow color and a white light color. They are color blended (as sources) onto the base image (the destination) –e.g. in OpenGL– using `glBlendFunc(GL_ZERO, GL_SRC_COLOR)`.

```

1:  for all viewpoints  $i$ 
2:    if  $i = 0$  then  $LC = \text{light color}$ 
3:    else  $LC = 1, 1, 1$ 
4:    set model-view-projection matrix to  $V_i$ 
5:    render real content into depth buffer
6:    render virtual content into stencil buffer
7:    render light in  $LC$  into frame buffer
   (previously cleared in 0,0,0)
8:    transfer frame buffer into texture
   memory  $T_i$ 
9:    categorize real content into fully visible
   ( $\Delta_{fi}$ ), partially visible ( $\Delta_{pi}$ ) and hidden
   ( $\Delta_{hi}$ ) triangles
10:   render  $\Delta_{fi}$  in  $LC$  into frame buffer (depth
   test disabled)
11:   transfer frame buffer into texture
   memory  $T_{i+\text{max\_vp}}$ 
12:  endfor
13:  set model-view-projection matrix to  $P$ , clear
   frame buffer in 0,0,0
14:  for all viewpoints  $i$ 
15:    if  $i \neq 0$  then enable color blending
16:    set texture matrix to  $V_i +$ 
   normalization space correction
17:    render  $\Delta_{fi}$  using projective texture  $T_i$ 
18:    render  $\Delta_{pi}$  using projective texture  $T_{i+\text{max\_vp}}$ 
19:    if  $i = 0$  then render  $\Delta_{hi}$  in light color
20:  endfor

```

Algorithm 2

In line 9, the real scene is categorized for each viewpoint into fully visible, partially visible and hidden triangles. These sets are rendered sequentially in lines 17-19. Fully visible triangles are texture mapped with the corresponding shadow mask and possibly color blended with the base image (line 17). Hidden triangles are rendered in the light color for the first viewpoint (to generate the base image), or are not rendered at all for all subsequent viewpoints (line 19). Triangles that are partially visible have to be partially texture mapped. To realize this without having to apply a time-consuming re-triangulation, a second shadow mask is generated for each

viewpoint (lines 10 and 11). Thereby, the original shadow mask is modified by rendering all fully visible triangles in the current light color on top of it (with the depth test disabled). Using this new shadow mask for texture mapping the partially visible triangles (line 18) ensures that the potential shadow area appears only on the visible portions of these triangles. The remaining part is then available for occlusion shadows of other viewpoints that can see these surface areas.

An efficient categorization of the triangles is achieved by storing and reusing the depth buffer that has been produced after line 5 is executed: Several sample points on a triangle are mapped from the world coordinate system into the screen coordinate system of the viewpoint. The calculated z-values of the transformed samples are then compared with the corresponding z-values, stored in the depth buffer. The depth buffer can be indexed using the computed x/y-coordinates of the transformed sample points. If all sample points have z-values that are closer to the viewpoint than the indexed values in the depth buffer, the triangle is categorized to be fully visible. If all points are further away than the indexed depth-buffer values, then the triangle is assumed to be hidden. If some points are closer and others are further away, then the triangle is partially visible.

Note that we currently apply these approximation by considering the three corner vertices and the center point of a triangle. All computations are cached and intermediate results are reused for triangles sharing the same vertices.

4.4. Multiple Projectors

Due to self occlusion, not all portions of the real content can be lit by a single light projector (e.g., surface 3 in figure 5). A solution to this problem is to increase the number of projectors and place them in such a way that the projected light is distributed over the real content. A set of optimal projector positions can be determined by applying Stuerzlinger's [13] hierarchical visibility algorithm. To guarantee a uniform illumination, however, surfaces should not be lit by more than one projector at the same time. Otherwise, the projected light accumulates on these surfaces and they appear brighter than others. Note that this effect is not necessarily spurious, since it reflects the natural behavior of multiple light sources (i.e., the light projectors) that illuminate the same surface. Consequently, we propose a solution to this problem that can be applied optionally.

Our method subdivides the geometry of the real content into surface portions that are assigned to, and finally rendered by an individual light projector. Since this subdivision is view-independent, and we assume that the parameters of the light projectors and the real content

do not change over time, it can be pre-computed. Algorithm 3 describes the off-line subdivision process.

Each triangle of the real content's geometry stores the following properties:

- a flag that indicates whether the triangle is fully visible (*visible*), completely hidden (*hidden*), or partially visible (*partial*) from a projector;
- the ID of the projector for which the *visible* flag applies (a triangle can be assigned to be fully visible by only one projector);
- a bit string with n bits for n projectors, indicating for which projectors the triangle is partially visible (the bit positions correspond to the projectors' IDs).

Note that a triangle's visibility from a particular view point differs from its visibility from a particular light projector. In this section, we describe only how to render triangles depending on their visibility from the projectors' perspective, while section 4.3 describes this with respect to the perspective of the view points. This should not lead to confusion.

```

1: initialize all triangles:  $\Delta = hidden$ 
2: for all projectors  $i$ 
3:   for all triangles  $j$ 
4:     if  $\Delta_j$  is fully visible from  $P_i$ 
5:        $A = Area(P_i, \Delta_j)$ 
6:       if  $\Delta_j = hidden$  or  $\Delta_j = partial$  or
7:         ( $\Delta_j = visible$  from  $P_k \{k < i\}$  and  $A_{\Delta_j} < A$ )
8:          $\Delta_j = visible$  from  $P_i$ 
9:          $A_{\Delta_j} = A$ 
10:      endif
11:    if  $\Delta_j$  is partially visible from  $P_i$  and  $\Delta_j \neq visible$ 
12:      then  $\Delta_j = partial$  from  $P_i$ 
13:    endif
14:  endfor

```

Algorithm 3

In line 1, all triangles are initialized to be *hidden* for all projectors. Every triangle (Δ_j) is then evaluated for each projector P_i . The algorithm assigns the following priority to the triangles: full visibility overwrites partial visibility, and partial visibility overwrites no visibility. Thus, a triangle will be assigned to be fully visible from the current projector (P_i) if (lines 4-10):

- it is fully visible from this projector *and*
- it has been previously assigned to be hidden or partially visible from another projector (P_k) or

- it has been previously assigned to be fully visible from another projector, but its projected area is larger from the current one.

Whether a triangle is completely hidden, partially visible, or fully visible from a specific perspective can be computed as described in section 4.3.

If a triangle is partially visible from the current projector and not fully visible by another one (lines 11-12), then it is assigned to be partially visible. In addition, the current projector is recorded in the triangle's bit string by activating the bit that corresponds to the projector's ID.

Finally, if a triangle is completely hidden from the current projector, nothing needs to be done and it remains hidden.

After all triangles have been assigned, a static shadow mask is generated for each projector. Therefore, only the fully visible triangles are rendered in a white light color on top of a black background – leaving the partially visible areas in black. The shadow masks are then read into separate blocks of the texture memory. Note that this is also part of the off-line pre-computation and has to be done only once. However, it is not explicitly outlined in algorithm 3.

During runtime, algorithm 1 (for a single view point) or algorithm 2 (for multiple viewpoints) are executed for each light projector separately (i.e., on different rendering hosts that are connected to a single light projector). The only modification to these algorithms is to restrict them to render only those triangles that have been assigned to the corresponding projector – not the entire real content. This affects only the underlying functionality of line 7 in algorithm 1, and lines 17-19 in algorithm 2. Note that a side effect of our approach is a distributed and potentially balanced rendering of the real content between different hosts.

In general, all assigned visible or partially visible triangles are rendered as described in algorithm 1 or algorithm 2. Partially visible triangles, however, require additional treatment: After being rendered for a particular projector in the discussed way (see sections 4.2 and 4.3), some of the static shadow masks that have been pre-computed for the other projectors are combined with the currently rendered image.

Technically, this is done exactly as for the multiple view points described in algorithm 2 – using projective texture mapping (but setting the texture matrix to P instead of V) and color blending.

If the blending function described in section 4.3 is used, white texture portions of these shadow maps do not effect the current image while black portions will erase the underlying image content. Consequently, previously lit portions are erased.

Specifically, the triangles' bit strings that indicate the set of projectors from which they are partially visible are evaluated. We define the following convention: Only the static shadow maps of those projectors whose IDs are smaller than the ID of the rendering projector have to be combined with the current image. Thus, we ensure that those portions of the partially visible triangles that have already been lit by a projector will be blocked for all other projectors.

Note that if a triangle is still marked as *hidden* after the subdivision, none of the projectors can illuminate it and it remains unlit. As long as the virtual light sources are located where the light projector is located, this case is treated properly. If the virtual light sources are located in arbitrary locations, these hidden triangles will potentially appear as incorrect shadow regions.

Figure 6 illustrates a simple example with two projectors (P1 and P2), one view point (V), and six real surfaces.

Using algorithm 3, we want to assume that surfaces 1 and 2 are assigned to be fully visible from P1, while surface 3, 5 and 6 are assigned to be fully visible from P2. Surface 4 is partially visible from both projectors. Consequently, surface 1 and 2 are only rendered from P1, and surfaces 3, 5 and 6 are only rendered from P2. P1 renders surface 4 first and illuminates portion *a*. Nothing else needs to be done for P1. Then P2 renders surface 4 and portion *b* is illuminated. Since projectors exist (i.e., only P1 in our case) that have previously lit a portion of surface 4⁵, the static shadow masks of these projectors are blended with the current image.

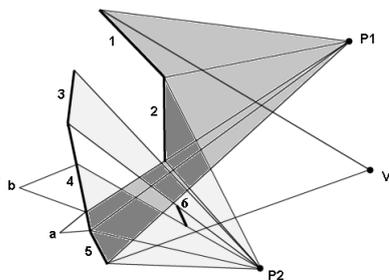


Figure 6: Visibility for different projectors.

In P1's static shadow mask, the image of portion *a* remains black while surfaces 1 and 2 are outlined in white. If mapped and blended into P2's image using the perspective texture transform of P1, portion *a* is erased in P2's image.

⁵ This is determined by comparing the projector IDs in the triangle's bit string with the ID of the current projector.

4.5. Drawing Light and Shadow

Surfaces of real objects for which the geometry is not known can be illuminated by manually drawing static light and shadow effects into the frame- and stencil-buffers of the corresponding light projectors. This allows the user to paint directly on these surfaces. In figures 1,3,4 and 7, for instance, the illumination of the base's lower part is static. It has been interactively sketched using a simple mouse-based drawing tool. More advanced painting techniques (e.g., as described by Bandyopadhyay et al. [2]) can be applied to support a more artistic expression.

5. Implementation

We use multiple off-the-shelf PCs with hardware-accelerated graphics boards that are connected via a standard local area network.

Each PC is connected to a single video projector and is executing the same render client. While one client drives the CRT projector that displays the stereoscopic images, an arbitrary number of other clients can be used to control the light projectors to render the illumination. All clients store an instance of the current scene in their local memories. If multiple projectors are used, each client pre-computes the entire subdivision of the real content's geometry and generates the static shadow masks for all projectors, as described in section 4.4.

A server that continuously receives user data from a tracking device represents the beginning of the chain. This data is passed to the first client that renders the stereo overlays. The client adds information about scene changes (e.g., caused by user interaction) to the tracking data and passes it to the next client in the chain. Based on this information, the client optionally adds additional information, passes it to the next client and renders the illumination. This is continued until the end of the chain is reached.

The synchronization between the render clients is realized in software and integrated into the communication protocol. Note that rendering and communication are carried out in parallel (i.e., within individual threads created by each client) – not causing synchronization delays.

For the experimental setup displayed in figure 2, two PCs and two projectors are used. The tracking server is running on the same PC as the first render client. The first projector is a standard CRT projector rendering the left and right eye views of the virtual objects in stereoscopic mode. For the optimal setup, the light projectors need to project independent occlusion shadows for the left and right eye in sync with the scene projection. This would require stereo genlock and frame locking between the master and client render processes and graphics systems.

For our experimental setup, we use a standard DLP projector in monoscopic mode as the light projector. The occlusion shadows are therefore generated only for an idealized viewpoint in-between the left and right eye position and we use the same occlusion shadows for both eyes. In our case, the stereo disparity for the virtual objects is rather small, since the virtual projection plane (i.e., the table top surface reflected into the Virtual Showcase) is in close proximity to the actual location of the virtual objects. This small stereo disparity results in a small disparity of the occlusion shadows. In addition, we typically defocus the light projector slightly to blur the occlusion shadow boundaries. In our experience, we found the monoscopic occlusion shadows to work quite well and that the disparity errors are hardly visible. In addition, the graphics cards are not genlocked, which results in slight flicker of the illuminated areas on the real objects.

6. Summary and Conclusions

We presented a solution for the problem of generating realistic occlusion effects with optical see-through displays. We describe an illumination technique that uses video projectors to generate view-dependent shadows underneath virtual overlays, which are projected directly onto the surface of physical objects. We have shown how this method can be applied for a single viewpoint and a single light projector. We also describe the necessary extensions for multiple users and multiple light projectors. In addition we discuss the limitations of our multi-user approach and suggest two techniques for reducing the artifacts.

The precision of the calibration method described in section 4.1 depends on the resolution of the frame-buffer that is displayed by the light projector, the distance between the projector and the reference surface and its orientation with respect to the projector's image plane. Surfaces that are aligned roughly parallel to the projector's image plane (e.g., the base in figures 1,3,4) allow us to achieve sub-pixel precision since the fiducials can be referenced exactly in the frame-buffer. Surfaces that are oriented more perpendicular to the image plane (e.g., skull in figure 7) still produce an average precision of 1-3 pixels⁶.

In addition, an object-based or image-based blurring can be applied to reduce the visual effects caused by slight registration errors, as described by Fuhrmann et al. [6]. Similar effects can be achieved by simply defocusing the projectors' optics – with no computational cost involved. We found that small displacements of the occlusion shadows are far less noticeable as

displacements between real content and virtual overlay. This might be due to the fact that one's focus is on the visible overlays and on the real objects, rather than on the occlusion shadows that are in general invisible.

The rendering algorithm for multiple viewpoints, described in section 4.3 can be efficiently applied in combination with real objects that physically divide the viewing space shared by multiple observers (such as the skull in figure 7). In these cases, the interference between different occlusion shadows can be minimized or even avoided completely since large portions of the surface exist that are not necessarily visible for more than one observer. Objects such as the wooden base in figures 1,3,4 on the other hand, consist of large surface portions that are likely to be visible from all viewpoints. In these cases, our second algorithm becomes less efficient since a categorization of the real scene's geometry is not necessary.

Simply blending all the shadow masks together and projecting them onto the real scene results in correct occlusion effects but also causes interference between different occlusion shadows. However, as mentioned in section 4.3, the occlusion shadows that can be perceived behave exactly like umbral hard-shadows that are cast by the virtual objects onto the real scene. By computing the matching illumination on the virtual objects' surfaces using point lights at the positions of the viewpoints amplifies this illusion and reduces the interference effect.

The algorithm for multiple projectors, described in section 4.4 avoids the problem that the same portion of the real content is illuminated by more than one projector. This is achieved by subdividing the real content's geometry and assigning the resulting portions to individual projectors. Raskar et al. [12], for instance, present a cross-feathering method to merge the images of multiple projectors on a pixel-basis, which could also be used with our setup.

7. Future Work

Our current prototype has proven that occlusion shadows strongly enhance the realistic display of real and virtual objects in a shared space. Occlusion shadows might also be of use for a variety of other displays that are based on the optical see-through concept. Head-attached displays including head-mounted displays (HMDs), and stationary displays such as projection-based AR devices [4] benefit from our method. However, a light controllable real environment is a prerequisite for displaying a high quality mixed reality scenario. Such an environment is implicitly provided by the Virtual Showcase.

For the Virtual Showcase, we are working on new optical elements and rendering techniques to reduce the number of light projectors necessary to reach almost all

⁶ With a frame-buffer resolution of 1024x768 pixels, and a distance between projector and reference surface of approx. 1.5m.

surfaces of real objects. Figure 8 illustrates a sketch of a potential optical configuration. We optically split up a single light frustum into multiple sub-frustums to provide a “surround illumination” with a single projector. However, this approach also splits the projector’s resolution.

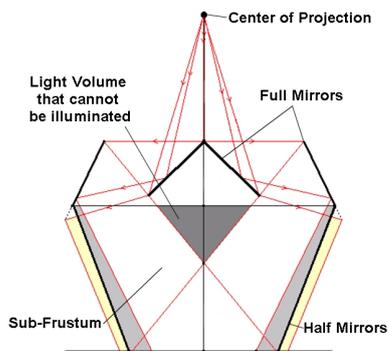


Figure 8: Sketch of a possible optical extension that splits a single light frustum into multiple sub-frustums providing a surround illumination with a single projector.

Our approach produces realistic occlusion effects between virtual and real objects. For the realistic display of mixed reality scenarios, consistent illumination of real and virtual objects is of great importance as well. As shown by Raskar et al. [12], real objects can be consistently illuminated with respect to virtual light sources using video projectors if their surface properties are known. We plan to apply this technique in the context of the Virtual Showcase to achieve a fully consistent and high quality display of mixed reality scenarios.

References

- [1] Azuma, R. T. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355-385, 1997.
- [2] Bandyopadhyay, D., Raskar, R., and Fuchs, H. Dynamic Shader Lamps: Painting on Real Objects. *In proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, pp. 207-215, 2001.
- [3] Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnaçao, L.M. The Virtual Showcase. *IEEE Computer Graphics & Applications*, vol. 21, no.6, pp. 48-55,2001.
- [4] Bimber, O., Encarnaçao, L.M. and Branco, P. The Extended Virtual Table: An Optical Extension for Table-Like Projection Systems. *Presence: Tele-operators and Virtual Environments*, vol.10, no. 6, 2001, pp. 613-631.
- [5] Breen, D.E., Whitaker, R. T., Rose, E. and Tuceryan, M. Interactive Occlusion and Automatic Object Placement for Augmented Reality. *Computer and Graphics Forum (proceedings of EUROGRAPHICS'96)*, vol. 15, no. 3, pp. C11-C22, 1996.
- [6] Fuhrmann, A., Hesina, G., Faure, F., and Gervautz, M. Occlusion in Collaborative Augmented Environments. *In proceedings of 5th Eurographics Workshop on Virtual Environments*, pp. 179-190, 1999.
- [7] Hua, H., Gao, C., Brown, L., Ahuja, N., and Rolland, J.P. Using a head-mounted projective display in interactive augmented environments. *In proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, pp. 217-223, 2001.
- [8] Kiyokawa, K., Kurata, Y. and Ohno, H. An Optical See-through Display for Mutual Occlusion of Real and Virtual Environments. *In proceedings of IEEE & ACM ISAR 2000*, pp. 60-67, 2000.
- [9] Naemura, T., Nitta, T., Mimura, A., Harashima, H. Virtual Shadows – Enhanced Interaction in Mixed Reality Environments. *In proceedings of IEEE Virtual Reality (IEEE VR'02)*, pp. 293-294, 2002.
- [10] Noda, S., Ban, Y., Sato, K., and Chihara, K. An Optical See-Through Mixed Reality Display with Realtime Rangefinder and an Active Pattern Light Source. *Transactions of the Virtual Reality Society of Japan*, vol. 4, no. 4, pp. 665-670, 1999.
- [11] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. Numerical Recipes in C - The Art of Scientific Computing (2nd edition), *Cambridge University Press, ISBN 0-521-43108-5*, pp. 412-420,1992.
- [12] Raskar, R. Welch, G., Low, K.L., and Bandyopadhyay, D. Shader Lamps: Animating real objects with image-based illumination. *In Proceedings of Eurographics Rendering Workshop (EGRW'01)*, 2001.
- [13] Stuerzlinger, W. Imagine all Visible Surfaces. *In Proceedings of Graphics Interface' 99*, pp. 115-122, 1999.

Acknowledgements

The Virtual Showcase project is supported by the European Union, IST-2001-28610.

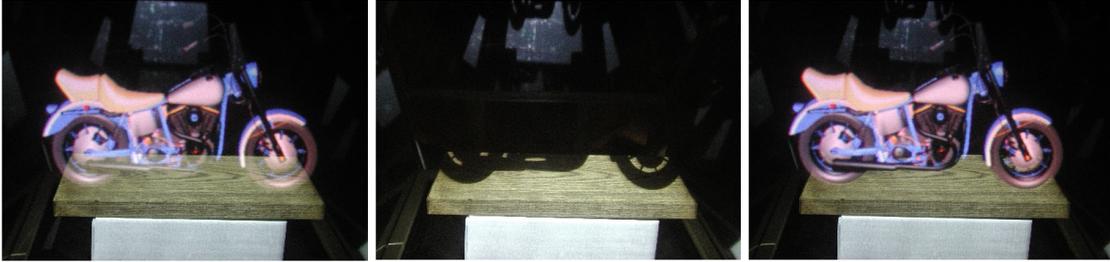


Figure 1: A normal illumination causes wrong occlusion effects between real objects and virtual overlays (left). Creating an occlusion shadow underneath the overlay prevents the light from being diffused on the real object's surface and transmitted through the optical combiner and the graphics (center). Overlaying the graphics over the shadow results in a realistic occlusion of the real object by the virtual one (left)⁷.



Figure 3: Knowing the depth information of the real content, also allows the occlusion of virtual objects by real ones (left). Combining this with our occlusion shadow method (center) creates correct mutual occlusion effects between both environments (right).



Figure 7: This example demonstrates our method in combination with a more complex real scene: A physical skull of a mid-cretaceous dinosaur (a *Deinonychus*) has been augmented with virtual muscles and bones (left) – generating occlusion shadows exactly underneath the virtual overlays. Covering the skull by a virtual skin (right) leaves most of the bone structure in shadow. Only the real teeth are clearly visible.

⁷ Note that the photographs in this paper have not been touched up. They were taken from the observer's point of view, but were rendered monoscopically.

Consistent Illumination within Optical See-Through Augmented Environments

Oliver Bimber, Anselm Grundhöfer, Gordon Wetzstein and Sebastian Knödel
Bauhaus University

Bauhausstraße 11, 99423 Weimar, Germany,

{oliver.bimber, anselm.grundhoefer, gordon.wetzstein, sebastian.knoedel}@medien.uni-weimar.de

Abstract

We present techniques which create a consistent illumination between real and virtual objects inside an application specific optical see-through display: the Virtual Showcase. We use projectors and cameras to capture reflectance information from diffuse real objects and to illuminate them under new synthetic lighting conditions. Matching direct and indirect lighting effects, such as shading, shadows, reflections and color bleeding can be approximated at interactive rates in such a controlled mixed environment.

1. Introduction

To achieve a consistent lighting situation between real and virtual environments is important for convincing augmented reality (AR) applications.

A rich pallet of algorithms and techniques have been developed that match illumination for video- or image-based augmented reality. However, very little work has been done in this area for optical see-through AR. For the reasons that are discussed in [1], we believe that the optical see-through concept is currently the most advanced technological approach to provide an acceptable level of realism and interactivity.

The Virtual Showcase [2] is an application specific optical see-through display. Figure 1 illustrates our latest prototype. It consists of up to four tilted CRT screens that are reflected by a pyramid-shaped mirror beam splitter. Wireless infrared tracking determines the observers' perspectives to render high-resolution¹ stereoscopic graphics onto the screens. Video projectors are mounted under its roof and allow a pixel-precise illumination of the real content [3]. Between two and three networked off-the-shelf PCs that are integrated into the Virtual Showcase's frame are used to drive the display. Beside its high resolution, the Virtual Showcase also provides a dark and well controllable environment that makes a consistent

illumination of real and virtual components easier to achieve than in real world environments.

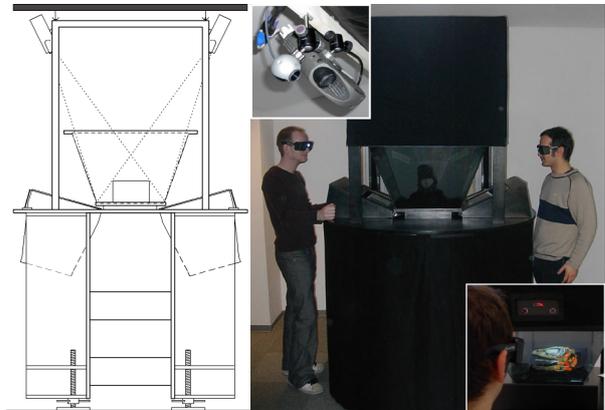


Figure 1: Virtual Showcase prototype with cameras and projectors.

The contribution of this paper is the introduction of methods which create a consistent illumination between real and virtual components within an optical see-through environment – such as the Virtual Showcase. Combinations of video projectors and cameras are applied to capture reflectance information from diffuse real objects and to illuminate them under new synthetic lighting conditions. For diffuse objects, the capturing process can also benefit from hardware acceleration – supporting dynamic update rates. To handle indirect lighting effects (like color bleeding) an off-line radiosity procedure is outlined that consists of multiple rendering passes. For direct lighting effects (such as simple shading, shadows and reflections) hardware accelerated techniques are described which allow to achieve interactive frame rates. The reflectance information is used in addition to solve a main problem of a previously introduced technique which creates consistent occlusion effects for multiple users within such environments [3].

¹ Currently UXGA per user.

2. Related Work

Inspired by the pioneering work of Nakamae et al. [16] and –later– Fournier et al. [9], many researchers have approached to create consistent illumination effects while integrating synthetic objects into a real environment. To our knowledge, all of these approaches represent the real environment in form of images or videos. Consequently, mainly image processing, inverse rendering, inverse global illumination, image-based and photo-realistic rendering techniques are applied to solve this problem. Due to the lack of real-time processing, these approaches are only applicable in combination with desktop screens and an unresponsive² user interaction. Devices that require interactive frame-rates, such as head-tracked personal or spatial displays, cannot be supported. Representative for the large body of literature that exists in this area, we want to discuss several more recent achievements:

Boivin et al. [5] present an interactive and hierarchical algorithm for reflectance recovery from a single image. They assume that the geometric model of the scenery and the lighting conditions within the image are known. Making assumptions about the scene’s photometric model, a virtual image is generated with global illumination techniques (i.e., ray-tracing and radiosity). This synthetic image is then compared to the photograph and a photometric error is estimated. If this error is too large, their algorithm will use a more complex BRDF model (step by step – using diffuse, specular, isotropic, and finally anisotropic terms) in the following iterations, until the deviation between synthetic image and photograph is satisfactory. Once the reflectance of the real scenery is recovered, virtual objects can be integrated and the scene must be re-rendered. They report that the analysis and re-rendering of the sample images takes between 30 minutes to several hours – depending on the quality required and the scene’s complexity.

Yu et al. [22] present a robust iterative approach that uses global illumination and inverse global illumination techniques. They estimate diffuse and specular reflectance, as well as radiance and irradiance from a sparse set of photographs and the given geometry model of the real scenery. Their method is applied to the insertion of virtual objects, the modification of illumination conditions and to the re-rendering of the scenery from novel viewpoints. As for Boivin’s approach, BRDF recovery and re-rendering are not supported at interactive frame-rates.

Loscos et al. [13] estimate only the diffuse reflectance from a set of photographs with different but controlled real world illumination conditions. They are able to insert and remove real and virtual objects and shadows, and to

modify the lighting conditions. To provide an interactive manipulation of the scenery, they separate the calculation of the direct and indirect illumination. While the direct illumination is computed on a per-pixel basis, indirect illumination is generated with a hierarchical radiosity system that is optimized for dynamic updates [8]. While the reflectance analysis is done during an offline preprocessing step, interactive frame-rates can be achieved during re-rendering. Depending on the performed task and the complexity of the scenery, they report re-rendering times for their examples between 1-3 seconds on a SGI R10000. Although these results are quite remarkable, the update rates are still too low to satisfy the high response requirements of stereoscopic displays that support head-tracking (and possibly multiple users).

Gibson and Murta [10] present another interactive image composition method to merge synthetic objects into a single background photograph of a real environment. A geometric model of the real scenery is also assumed to be known. In contrast to the techniques described above, their approach does not consider global illumination effects to benefit from hardware accelerated multi-pass rendering. Consequently, a reflectance analysis of the real surfaces is not required, indirect illumination effects are ignored, and a modification of the lighting conditions is not possible. The illumination of the real environment is first captured in form of an omnidirectional image. Then a series of high dynamic basis radiance maps are pre-computed. They are used during runtime to simulate a matching direct illumination of the synthetic objects using sphere mapping. Shadow casting between real and virtual objects is approximated with standard shadow mapping. With their method, convincing images can be rendered at frame rates up to 10fps on an SGI Onyx 2. However, it is restricted to a static viewpoint.

3. Diffuse Reflectance Analysis

We want to assume that the geometry of both object types –real and virtual– has been modeled or scanned in advance. While the material properties of virtual objects are also defined during their modeling process, the diffuse reflectance of physical objects is captured inside the Virtual Showcase with a set of video projectors and cameras. This sort of analysis is standard practice for many range scanner setups. But since we consider only diffuse real objects (a projector-based illumination will generally fail for any specular surface), our method can benefit from hardware accelerated rendering techniques. In contrast to conventional scanning approaches, this leads to dynamic update rates. Figure 2 illustrates an example.

² Not real-time.

3.1. Calibration

The intrinsic and extrinsic parameters of projectors and cameras within the world coordinate system have to be estimated first. We calibrate each device separately. As described in [3], we interactively mark the two-dimensional projections of known three-dimensional fiducials on a projector's/camera's image plane. Using these mappings, we apply Powell's direction set method [18] to solve a perspective n-point problem for each device. This results in the perspective projection matrices P, C of a projector and a camera. Both matrices incorporate the correct model-view transformations with respect to the origin of our world coordinate system.

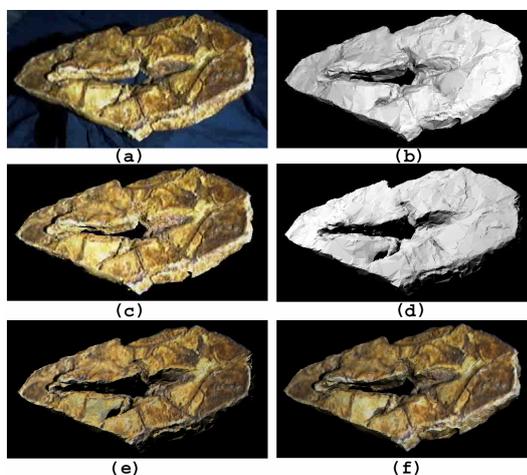


Figure 2: (a) Captured radiance map of a fossilized dinosaur footprint; (b) Intensity image rendered for calibrated projector from (a); (c) Computed reflectance map; (d) Novel illumination situation; (e) Reflectance map under novel illumination from (d); (f) Reflectance map under virtual illumination from (b).

3.2. Capturing Radiance Maps

A video projector is used to send structured light samples to the diffuse physical object and illuminate it with a predefined color C_p and an estimated intensity η . Synchronized to the illumination, a video camera captures an input image. Since this image contains the diffuse reflectance of the object's surface under known lighting conditions it represents a *radiance map*. White-balancing and other dynamic correction functions have been disabled in advance. The parameters of the camera's response function are adjusted manually in such a way that the recorded images approximate the real world situation as close as possible.

Some types of video projectors (such as digital light projectors, DLPs) display a single image within sequential, time-multiplexed light intervals to achieve different intensity levels per color. If such projectors are used, a single snapshot of the illuminated scene would capture only a slice of the entire display period. Consequently, this image would contain incomplete color fragments instead of a full-color image. The width of this slice depends on the exposure time of the camera. To overcome this problem, and to be independent of the camera's exposure capabilities, we capture a sequence of images over a predefined period of time. These images are then combined and averaged to create the final diffuse radiance map I_{rad} (cf. figure 2a).

3.3. Creating Intensity Images

To extract the diffuse material reflectance out of I_{rad} the lighting conditions that have been created by the projector have to be neutralized. OpenGL's diffuse lighting component is given by [17]:

$$I_i = \frac{1}{r_i^2} \cos(\theta_i) (D_l D_m)_i \quad (3.1)$$

where I_i is the final intensity (color) of a vertex i , D_l is the diffuse color of the light, D_m is the diffuse material property, the angle θ_i is spanned by the vertex's normal and the direction vector to the light source, and the factor $1/r_i^2$ represents a square distance attenuation.

Similar as in [19], an intensity image I_{int} that contains only the diffuse illumination can be created by rendering the object's geometry (with $D_m = 1$) from the perspective of the video camera, illuminated by a point light source (with $D_l = C_p \eta$) that is virtually located at the position of the projector (cf. figure 2b).

In addition, hard shadows are added to the intensity image by applying standard shadow mapping techniques. Consequently, the background pixels of I_{int} , as well as pixels of regions that are occluded from the perspective of the light source are blanked out ($I_{int}(x, y) = 0$), while all other pixels are shaded under consideration of equation 3.1.

3.4. Extracting and Re-Rendering Diffuse Reflectance

Given the captured radiance map I_{rad} and the rendered intensity image I_{int} , the diffuse reflectance for each surface that is visible to the camera can be computed by:

$$I_{ref}(x, y) = \begin{cases} \frac{I_{rad}(x, y)}{I_{int}(x, y)} & \text{for } I_{int}(x, y) > 0, \\ 0 & \text{for } I_{int}(x, y) = 0 \end{cases} \quad (3.2)$$

We store the reflectance image I_{ref} , together with the matrix C and the real object's world-transformation O_c that is active during the capturing process within the same data-structure. We call this data structure *reflectance map* (cf. figure 2c).

The captured reflectance map can be re-rendered together with the real object's geometric representation from any perspective with an arbitrary world-transformation O_a . Thereby, I_{ref} is applied as projective texture map with the texture matrix³ set to $O_a^{-1}O_cC$. Enabling texture modulation, it can then be re-lit virtually under novel illumination conditions (cf. figures 2d, 2e and 2f).

3.5. Shortcomings and Solutions

The basic reflectance analysis method as described above faces the following problems:

- (1) due to under-sampling, surfaces which span a large angle ϕ_i between their normal vectors and the direction vectors to the camera can cause texture artifacts if I_{ref} is re-mapped from a different perspective;
- (2) a single reflectance map covers only the surface portion that is visible from the perspective of the camera;
- (3) the radiance map can contain indirect illumination effects caused by light fractions that are diffused off other surfaces (so-called *secondary scattering*). The intensity image I_{int} , however, does not contain secondary scattering effects since a global illumination solution is not computed. Consequently, the extracted reflectance is incorrect at those areas that are indirectly illuminated by secondary scattering;
- (4) the projector intensity η has to be estimated correctly;

To overcome the under-sampling problem, we define that only surfaces with $\phi_i \leq \phi_{max}$ are analyzed. All other surfaces will be blanked-out in I_{ref} (i.e., $I_{ref}(x, y) = 0$). We found that $\phi_{max} = 60^\circ$ is appropriate. This corresponds

³ Including the corresponding mapping transformation from normalized device coordinates to normalized texture coordinates.

to the findings in [19], describing a maximum angle between projector and projection surface.

Multiple reflectance maps that cover different surface portions can be captured under varying transformations O_c or C . They are merged and alpha blended during re-mapping them via multi-texturing onto the object's geometric representation. This ensures that regions which are blanked out in one reflectance map can be covered by other reflectance maps. To generate seamless transitions between the different texture maps, bi- or tri-linear texture filtering can be enabled.

Illuminating the entire scene can cause an extreme secondary scattering of the light. To minimize the appearance of secondary scattering in I_{rad} , we divide the scene into discrete pieces and capture their reflectance one after the other. For this, we can apply the same algorithm as described above. The difference, however, is that we illuminate and render only one piece at a time which then appears in I_{rad} and I_{int} . By evaluating the blanked out background information provided in I_{int} , we can effectively segment the selected piece in I_{int} and compute its reflectance. This is repeated for each front-facing piece, until I_{ref} is complete.

We estimate the projector's intensity η as follows: First, we generate a reflectance map with an initial guess of η . This reflectance map is then re-mapped onto the object's geometric representation, which is rendered from the perspective of the camera and illuminated by a virtual point light source with η located at the projector. The rendered radiance map I_{radv} is then compared to the captured radiance map I_{rad} by determining the average square distance error Δ among all corresponding pixels. Finally, we find an approximation for η by minimizing the error function f_Δ . For this we apply Brent's inverse parabolic minimization method with bracketing [7]. By estimating η , we can also incorporate the constant black-level of the projector.

4. Augmented Radiosity

In computer graphics, the radiosity method [11] is used to approximate a global illumination solution by solving an energy-flow equation. Indirect illumination effects, such as secondary scattering can be simulated with radiosity. The general radiosity equation for n surface patches is given by:

$$B_i = E_i + \rho_i \sum_{j=1}^n B_j F_{ij} \quad (4.1)$$

where B_i is the radiance of surface i , E_i is the emitted energy per unit area of surface i , ρ_i is the reflectance of

surface i , and F_{ij} represents the fraction of energy that is exchanged between surface i and surface j (the form-factor).

The simulation of radiosity effects within an optical see-through environment that consists of diffuse physical and virtual objects, is facing the following challenges and problems:

- (1) light energy has to flow between all surfaces – real ones and virtual ones;
- (2) physical objects are illuminated with physical light sources (i.e., video projectors in our case) which do not share the geometric and radiometric properties of the virtual light sources;
- (3) no physical light energy flows from virtual objects to real ones (and vice versa). Consequently, the illuminated physical environment causes (due to the absence of the virtual objects) a different radiometric behavior than the entire environment (i.e., real and virtual objects together).

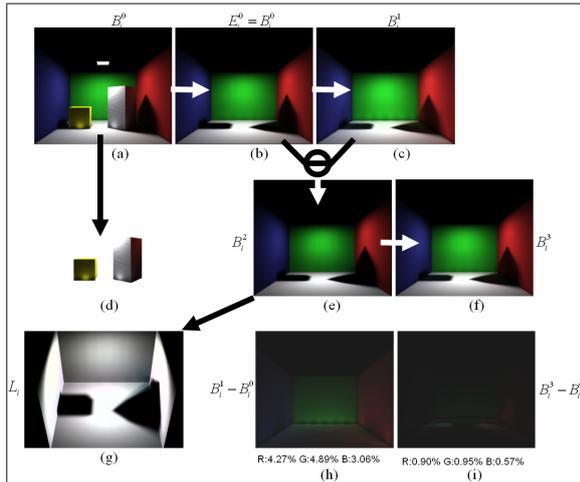


Figure 3: Multi-Pass Radiosity.

An example is illustrated in figure 3a⁴. The entire environment consists of three walls, a floor, two boxes and a surface light source on the ceiling. We want to assume that the walls and the floor are the geometric representations of the physical environment, and the boxes as well as the light source belong to the virtual environment. While the diffuse reflectance ρ_i of the physical environment can be automatically captured (as described in section 3), it has to be defined for the virtual environment. After a radiosity simulation⁵ of the entire

⁴ We have chosen a physical mock-up of the Cornell room since it is used in many other publications as a reference to evaluate radiosity techniques.

⁵ We applied a hemi-cube-based radiosity implementation with progressive refinement, adaptive subdivision and interpolated rendering for our simulations.

environment the radiance values B_i^0 for all surfaces have been computed⁶. Color-bleeding and shadow-casting are clearly visible.

4.1 Virtual Objects

For virtual objects, the computed radiance values are already correct (cf. figure 3d). The rendered image represents a radiance map that is generated from one specific perspective. Rendering the virtual objects from multiple perspectives results in multiple radiance maps that can be merged and alpha blended during re-mapping them via multi-texturing onto the virtual geometry (as described for reflectance maps in section 3.5). In this case, our radiance maps are equivalent to light maps that are often being applied during pre-lighting steps to speed up the online rendering process.

The pre-lit virtual objects can simply be rendered together with their light maps and can be optically overlaid over the physical environment.

4.2 Physical Objects

The physical surfaces, however, have to emit the energy that was computed in B_i^0 (cf. figure 3b). To approximate this, we first assume that every physical surface patch directly emits an energy E_i^0 that is equivalent to B_i^0 . If this is the case, fractions of this energy will radiate to other surfaces and illuminate them in addition. This can be simulated by a second radiosity-pass (cf. figure 3c), which computes new reflectance values B_i^1 for all the physical surfaces, by assuming that $E_i^0 = B_i^0$, and not considering the direct influence of the virtual light source.

If we subtract the radiance values that have been computed in both passes we receive the scattered light only. That is, the light energy radiated between the physical surfaces $B_i^1 - B_i^0$ (cf. figure 3h).

Consequently,

$$B_i^2 = B_i^0 - (B_i^1 - B_i^0) \quad (4.2)$$

approximates the energy that has to be created physically on every real surface patch. To prove this we can apply a third radiosity pass to simulate the energy flow between the patches (cf. figure 3f). We can see that the remaining energy $B_i^1 - B_i^0$ will be nearly added, and we receive:

$$B_i^3 = B_i^2 + (B_i^1 - B_i^0) \approx B_i^0 \quad (4.3)$$

By removing the virtual objects from the environment and simulating the second radiosity pass, light energy will also be radiated onto surfaces which were originally blocked or covered by the virtual objects (either

⁶ Note, that the upper index represents the radiosity pass.

completely or partially). Examples are the shadow areas that have been cast by the virtual objects. This can be observed in figure 3h and figure 3i. Consequently, negative radiance values are possible for such areas after applying equation 4.2. To avoid this, the resulting values have to be clipped to a valid range.

The average deviations between B_i^0 and B_i^1 , as well as between B_i^0 and B_i^3 , within the three spectral samples red (R), green (G), and blue (B) are presented below figures 3h and 3i, respectively. Treating a video projector as a point light source B_i^2 can be expressed as a simplified version of equation 4.1:

$$B_i^2 = \rho_i L_i F_i \quad (4.4)$$

where L_i is the irradiance that has to be projected onto surface i by the projector, and F_i is the form-factor for surface i , which is given by:

$$F_i = \frac{\cos(\theta_i)}{r_i^2} h_i \quad (4.5)$$

where θ_i is the angle between a surface patch's normal and the direction vector to the projector, r_i is the distance between a surface patch and the projector, and h_i is the visibility term of the surface patch, seen from the projector's perspective.

Extending and solving equation 4.4 for L_i , we receive (cf. figure 3g):

$$L_i = \frac{B_i^2}{\rho_i F_i} \eta \quad (4.6)$$

To cope with the individual brightness of a video projector, we add the intensity factor η . How to estimate η for a specific projector was described in section 3.5. To be consistent with our previously used terminology, we call L_i the *irradiance map*.

4.3 Limitations

The computed radiance and irradiance values are view-independent. Consequently, irradiance maps for the real objects and radiance maps for the virtual objects can be pre-computed offline.

The real objects are illuminated with projected light during runtime by rendering the generated irradiance map from the viewpoint of the projector (e.g., as illustrated in figure 3g). Virtual objects are rendered with the computed light maps (e.g., as illustrated in figure 3d) and are then optically overlaid over the real environment. Due to the view-independence of the method, the augmented scene can be observed from any perspective (i.e., head-tracking and stereoscopic rendering are possible). However, the scene has to remain static, since any modification would require to re-compute new radiance and irradiance maps

throughout multiple radiosity passes. This is not yet possible at interactive rates.

Figure 4 shows a photograph of (a) the physical object under room illumination, (b) a screen-shot of captured reflectance maps that have been re-rendered under novel lighting conditions, (c) a screen-shot of the simulated radiance situation B_i^0 , and (d) a photograph of a physical object that has been illuminated with L_i . Note, that small deviations between the images can be contributed to the responds of the digital camera that was used to take the photograph, as well as to the high black-level of the projector that, for instance, makes it impossible to create completely black shadow areas.

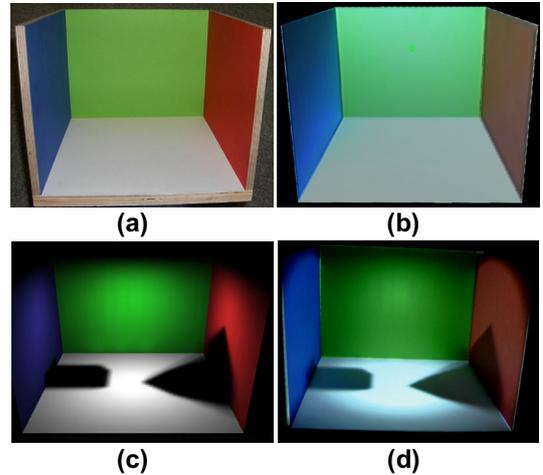


Figure 4: (a) Photograph of original object under room illumination; (b) Screen-shot of captured reflectance re-lit with virtual point light source and Phong shading; (c) Screen-shot of simulated radiosity solution with captured reflectance, virtual surface light source (shown in figure 3), and two virtual objects (shown in figure 3); (d) Photograph of original object illuminated with the computed irradiance.

5. Interactive Approximations

In the following section we describe several interactive rendering methods that make use of hardware acceleration. In particular we discuss how to create matching shading, shadow and reflection effects on real and virtual objects. Indirect lighting effects such as color-bleeding, however, cannot be created with these techniques. Yet, they create acceptable results at interactive frame rates for multiple head-tracked users and stereoscopic viewing on conventional PCs.

5.1 Shading

The generation of direct illumination effects on virtual surfaces caused by virtual light sources is a standard task of today's hardware accelerated computer graphics technology. Real-time algorithms, such as Gouraud shading or Phong shading are often implemented on graphics boards.

Consistent and matching shading effects on real surfaces from virtual light sources can be achieved by using video projectors that project appropriate irradiance maps onto the real objects. Raskar et al. [19] show how to compute an irradiance map to lift the radiance properties of neutral diffuse objects with uniform white surfaces into a pre-computed radiance map of a virtual scene illuminated by virtual light sources. An irradiance map that creates virtual illumination effects on diffuse real objects with arbitrary reflectance properties (color and texture) can be computed as follows:

First, the real objects' captured reflectance map (I_{ref}) is rendered from the viewpoint of the projector and is shaded with all virtual light sources in the scene. This results in the radiance map I_{rad_1} . Then I_{ref} is rendered again from the viewpoint of the projector. This time, however, it is illuminated by a single point light source (with $D_i = 1 \cdot \eta$) which is located at the position of the projector. This results in the radiance map I_{rad_2} . Finally, the correct irradiance map is computed by:

$$L = \frac{I_{rad_1}}{I_{rad_2}} \quad (5.1)$$

Note that equation 5.1 correlates to equation 4.6. The difference is the applied illumination model. While equation 4.6 is discussed with respect to an indirect global illumination model (radiosity), equation 5.1 applies hardware accelerated direct models (such as Phong or Gouraud shading). It is easy to see that I_{rad_1} is the opponent to B_i^2 and that I_{rad_2} corresponds to $\rho_i F \eta$.

Note also that this method is actually completely independent of the real objects' reflectance. This can be shown by equalizing I_{rad_1} with I_{rad_2} through equation 3.1. In this case the diffuse material property D_m (i.e., the reflectance) is canceled out. Consequently, I_{rad_1} and I_{rad_2} can be rendered with a constant (but equal) reflectance (D_m). If we choose $D_m = 1$ then the irradiance map is simply the quotient between the two intensity images I_{int_1} and I_{int_2} that result from the two different lighting conditions – the virtual one and the real one.

The irradiance map L should also contain consistent shadow information. How to achieve this is outlined

below. Figure 5 illustrates examples with matching shading effects⁷.

5.2 Shadows

We can identify six types of shadows within an optical see-through environment:

- (1) shadows on real objects created by real objects and real light sources;
- (2) shadows on virtual objects created by virtual objects and virtual light sources;
- (3) shadows on virtual objects created by real objects and virtual light sources;
- (4) shadows on real objects created by real objects and virtual light sources;
- (5) shadows on real objects created by virtual objects and virtual light sources;
- (6) occlusion shadows;

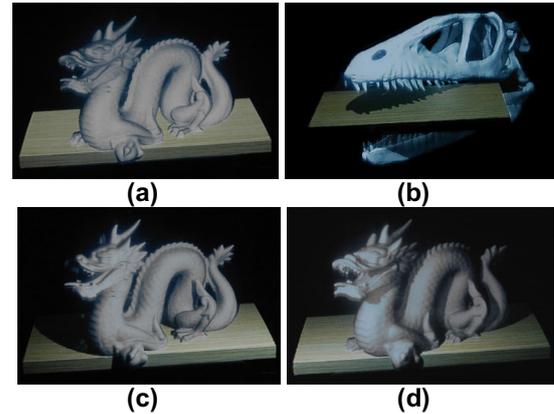


Figure 5: (a) Unrealistic uniform illumination with shadow type 6 (the wooden plate is real, the dragon and the dinosaur skull are virtual); (b)-(d) Realistic illumination under varying virtual lighting conditions with matching shading and shadows (types 2,3,5, and 6).

The first type of shadow is the result of occlusions and self-occlusions of the physical environment that is illuminated by a physical light source (e.g., a video projector). Since we focus on controlling the illumination conditions within the entire environment via virtual light sources we have to remove these shadows. This can be achieved by using multiple synchronized projectors that are able to illuminate all visible real surfaces. Several techniques have been described which compute a correct color and intensity blending for multi-projector displays [14, 19, 21].

⁷ Note that we have chosen a simple wooden plate to demonstrate and to compare the different effects. However, all techniques that are explained in this paper can be applied to arbitrary object shapes.

The second and third shadow types can be created with standard shadow mapping or shadow buffering techniques. To cast shadows from real objects onto virtual ones, the registered geometric representations of the real objects have to be rendered together with the virtual objects when the shadow map is created (i.e., during the first shadow pass). Such geometric real world representations (sometimes called *phantoms* [6]) are often rendered continuously to generate a realistic occlusion of virtual objects by real ones. Note that these hardware accelerated techniques create hard shadows while global illumination methods (such as radiosity) can create soft shadows. Texture blending, however, allows ambient light to be added to the shadow regions. This results in dark shadow regions that are blended with the underlying surface texture, instead of creating unrealistic black shadows.

Shadow types number 4 and 5 can also be created via shadow mapping. However, they are projected on the surface of the real object together with the irradiance map L , as discussed in section 5.1. Therefore, I_{rad_1} has to contain the black (non-blended) shadows of the virtual and the real objects. This is achieved by rendering all virtual objects and all phantoms during the first shadow pass to create a shadow map. During the second pass the shaded reflectance texture and the generated shadow texture are blended and mapped onto the objects' phantoms. A division of the black shadow regions by I_{rad_2} preserves these regions. Note that a blending of the projected shadows with the texture of the real objects occurs physically if the corresponding surface portions are illuminated (e.g., by a relatively small amount of projected ambient light).

Occlusion shadows [3] are special view-dependent shadows created by the projectors on the real objects' surfaces. We have introduced them to achieve a realistic occlusion of real objects by virtual ones. They are normally not visible from the perspective of the observer, since they are displayed exactly underneath the graphical overlays. Occlusion shadow-maps, however, have also to be blended to the irradiance map L before it is projected.

5.3 Reflections

Using hardware accelerated cube mapping techniques, the virtual representation of the real environment (i.e., the objects' geometry together with the correctly illuminated reflectance map) can be reflected by virtual objects (cf. figure 6). Therefore, only the registered virtual representation of the real environment has to be rendered during the generation step of the cube map. Virtual objects are then simply rendered with cube mapping enabled. Note, that for conventional cube mapping, reflection effects on a virtual object are physically correct

for only a single point – the center of the cube map frusta. To create convincing approximations this center has to be matched with the virtual object's center of gravity, and the cube map has to be updated every time the scene changes.

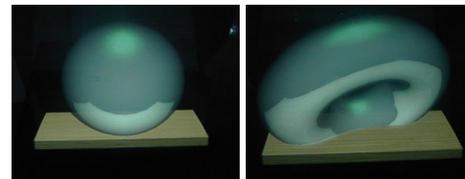


Figure 6: (a) A virtual sphere and (b) a virtual torus reflecting and occluding the real object (wooden plate).

5.4 Occluding Occlusion Shadows

The occlusion shadow method [3] is currently one of two functioning solutions that can create consistent occlusions effects for optical see-through displays. A main drawback of this approach is its limited support for multiple users: If the same real surfaces are simultaneously visible from multiple points of view (as it is the case for different observers), individual occlusion shadows that project onto these surfaces are also visible from different viewpoints at the same time (cf. figure 7a).

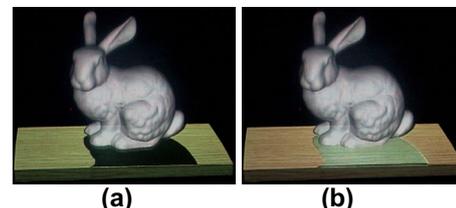


Figure 7: (a) Occlusion shadow of second observer is clearly visible; (b) Wrongly visible occlusion shadow is covered by optically overlaying the corresponding part of the reflectance map.

Although two different approaches have been presented in [3] that reduce the effects of this problem, it is not completely solved for any type of surface. Knowing the reflectance information of the real surfaces, however, leads to an effective and general solution:

As described in sections 5.1 and 5.2, the real objects are illuminated by projected light (also containing occlusion shadows for all observers) and the virtual objects are shaded, rendered and optically overlaid over the real scene (on top of each observer's occlusion shadows). In addition to the virtual scene, we render the portions of the real scene (i.e., its registered reflectance

map) that are covered by the occlusion shadows of all other observers.

Remember that these reflectance-map portions are illuminated and shaded under the same lighting conditions as their real counterparts (outlined in sections 5.1 and 5.2). This creates seamless transitions between the real and the virtual parts.

For each observer the occlusion shadows of all other observers are rendered into the stencil buffer first. This is done by rendering the real scene's geometry from each observer's perspective and adding the corresponding occlusion shadows via projective texture mapping, as described in [3]. The stencil buffer has to be filled in such a way that the area surrounding the occlusion shadows will be blanked out in the final image. Then the real scene's reflectance map is rendered into the frame buffer (also from the perspective of the observer) and is shaded under the virtual lighting situation. After stenciling has been disabled, the virtual objects can be added to the observer's view (cf. figure 7b).

6. Summary and Future Work

The overall goal of this work is to enhance realism for optical see-through AR environments. To achieve this we have developed techniques which allow creating consistent illumination effects between real and virtual objects. We have implemented and demonstrated these techniques based on the Virtual Showcase, since this display provides a well controllable environment.

We use video projectors and cameras as essential components of the Virtual Showcase. They allow to retrieve information out of the Virtual Showcase's inside at dynamic update rates, and to illuminate real objects on a per-pixel basis in real time. Currently only reflectance data is scanned from real objects. In the future, other surface information, such as geometry and external emission can be measured as well. The scanned geometry information will lead to the development of an automatic registration procedure for real objects.

Using the reflectance information, Augmented Radiosity has been described as a global illumination technique for optical see-through devices that is able to create a high level of realism. Only static augmented environments can be created with this method. However, due to its view-independency, head-tracking and stereoscopic rendering is possible. To evaluate interactive global illumination methods in combination with this technique belongs to our list of future tasks.

To reach interactive rendering frame rates, hardware accelerated methods have been used to generate convincing approximations of a consistently and realistically illuminated augmented environment. Throughout several rendering passes shading, shadow mapping and cube mapping techniques have been applied

in combination with the captured reflectance information to achieve this. Instead of capturing the reflectance map of a real object and computing a radiance map, its physical radiance can be captured directly after the synthetic illumination information (i.e., shading and shadows, as described in section 5.1) have been created on its surface. This radiance map can then be applied in combination with the reflection and occlusion shadow techniques described in sections 5.3 and 5.4. The advantage of this approach is that the illumination information on the real object has to be computed only once – before it is projected onto its surface. The resulting radiance map can simply be captured with the video camera and contains all important information of the real environment, such as reflectance, shading and shadows.

The occluding occlusion shadow technique allows to solve the multi-user limitation that is linked to the original occlusion shadow idea. In addition, it allows to make seamless transitions on the mixed reality continuum [15], that are important for applications of the Virtual Showcase, such as digital storytelling [4]. The biggest problem of this extension, however, is a slight color inconsistency between the real object and the virtual overlay. This results from photometric deviations between the CRT screens and the video cameras with the observers' visual perception of the real object. Currently we adjust this manually by modifying the physical and the synthetic illumination until the real object and the virtual overlay appear to coincide visually. Automatic calibration techniques need to be developed in the future. To reduce geometric misalignments caused by small registration errors, the edges of the virtual overlay can be blurred in addition.

A next important step towards realism will be to visually enhance the virtual components while retaining interactive frame rates. Advanced rendering techniques, such as light fields [12, 20] and hardware-accelerated procedural shading technology might allow blurring the boundaries between real and virtual even further.

Acknowledgements

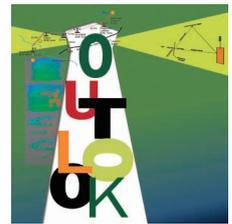
We thank Werner Purgathofer for fruitful discussions on the occluding occlusion shadows topic. The Virtual Showcase project is supported by the European Union, IST-2001-28610.

References

- [1] Azuma, R., Baillet, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B. Recent Advances in Augmented Reality. *IEEE Computer Graphics & Applications*, vol. 21, no.6, pp. 34-47, 2001.

- [2] Bimber, O., Fröhlich, B., Schmalstieg, D., and Encarnação, L.M. The Virtual Showcase. *IEEE Computer Graphics & Applications*, vol. 21, no.6, pp. 48-55, 2001.
- [3] Bimber, O., and Fröhlich, B. Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays. In proceedings of ACM/IEEE International Symposium on Mixed and Augmented Reality (ISMAR'02), pp. 186-195, 2002.
- [4] Bimber, O., Encarnação, L.M, and Schmalstieg, D. The Virtual Showcase as a new Platform for Augmented Reality Digital Storytelling. In proceedings of IPT/EGVE 2003 workshop, pp. 87-95, 2003.
- [5] Boivin, S. and Gagalowicz, A. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In proceedings of ACM Siggraph, pp. 107-116, 2001.
- [6] Breen, D.E., Whitaker, R. T., Rose, E. and Tuceryan, M. Interactive Occlusion and Automatic Object Placement for Augmented Reality. *Computer and Graphics Forum* (proceedings of EUROGRAPHICS'96), vol. 15, no. 3, pp. C11-C22, 1996.
- [7] Brent, R.P. Algorithms for minimization without derivatives. Prentice-Hall, Engelwood Cliffs, NJ, 1973.
- [8] Drettakis, G. and Sillion, F. Interactive update of global illumination using a line-space hierarchy. In proceedings of ACM Siggraph'97, pp- 57-64, 1997.
- [9] Fournier, A. Gunawan, A.S., and Romanzin, C. Common illumination between real and computer generated scenes. In proceedings of Graphics Interface'93, pp. 254-262, 1993.
- [10] Gibson, S. and Murta, A. Interactive rendering with real-world illumination. In proceedings of 11th Eurographics Workshop on Rendering, pp. 365-376, 2000.
- [11] Goral, C.M., Torrance, K.E., Greenberg, D.P., and Battaile, B. Modeling the interaction of light between diffuse surfaces. *Computer In proceedings of ACM Siggraph'84*, vol. 18, no. 3, pp. 212-222, 1984.
- [12] Levoy, M. and Hanrahan, P. Light field rendering. In proceedings of ACM Siggraph'96, pp. 31-42, 1996.
- [13] Loscos, C., Drettakis, G., and Robert, L. Interactive virtual relighting of real scenes. *IEEE Transactions on Visualization and Computer Graphics*, vol, 6, no. 3, pp. 289-305, 2000.
- [14] Majumder, A. He, Z., Towles, H., and Welch, G. Achieving color uniformity across multi-projector displays. In proceedings of IEEE Visualization, pp. 117-124, 2000.
- [15] Milgram, P. and Kishino, F. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems E77-D*, vol.12, pp. 1321-1329, 1994.
- [16] Nakamae, E., Harada, K., Ishizaki, T., and Nishita, T. A montage method: The overlaying of computer generated images onto background photographs. In proceedings of ACM Siggraph'86, pp. 207-214, 1986.
- [17] Neider, J., Davis, T., and Woo, M. *OpenGL Programming Guide. Release 1*, Addison Wesley, ISBN 0-201-63274-8, pp.157-194, 1996.
- [18] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. *Numerical Recipes in C - The Art of Scientific Computing* (2nd edition), Cambridge University Press, ISBN 0-521-43108-5, pp. 412-420, 1992.
- [19] Raskar, R. Welch, G., Low, K.L., and Bandyopadhyay, D. Shader Lamps: Animating real objects with image-based illumination. In Proc. of Eurographics Rendering Workshop (EGRW'01), 2001.
- [20] Wood, D.N., Azuma, D.I., Aldinger, K., Curless, B., Duchamp, T., Salesin, D.H., and Stuetzle, W. Surface Light Fields for 3D Photography, In proceedings of ACM Siggraph'00, pp 287-296, 2000.
- [21] Yang, R., Gotz, D., Hensley, J., Towles, H., and Brown, M.S. PixelFlex: A reconfigurable multi-projector display system. In proceedings of IEEE Visualization, pp. 167-174, 2001.
- [22] Yu, Y., Debevec, P., Malik, J., and Hawkins, T. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In proceedings of ACM Siggraph' 99, pp. 215-224, 1999.

Combining Optical Holograms with Interactive Computer Graphics



Merging optical holograms with 3D graphical elements can provide an acceptable tradeoff between quality and interactivity: The holographic data provides high-quality but static content, while additional graphical information can be generated, inserted, modified, and animated at interactive rates.

*Oliver
Bimber*
Bauhaus University

Today, many applications for optical holograms exist, including interferometry, copy protection, data storage, and holographic optical elements. A hologram is a photometric emulsion that records interference patterns of coherent light. The recording itself stores the amplitude, wavelength, and phase information of light waves. In contrast to simple photographs, which can record only amplitude and wavelength information, holograms can reconstruct complete optical wavefronts. This results in the captured scenery having a three-dimensional appearance that can be observed from different perspectives.

Museum exhibits often use optical holograms because they can present 3D objects with almost no loss in visual quality. Displaying artifacts virtually removes the need to build physical replicas of the original objects. In addition, this holographic technology can be used to make medical, dental, archaeological, and other recordings—both for teaching and documentation.

Optical holograms are static, however, and lack interactivity. Multiplex holograms offer an apparent exception. Built from multiple vertical-strip holograms that contain recordings of the same scenery at different time intervals, they let observers perceive the recorded scene in motion while moving around the hologram or by spinning a cylindrically

shaped version around its principal axis. Multiplex holograms are not, however, truly interactive.

Combining 3D computer graphics with stereoscopic presentation techniques provides an alternative that allows interactivity. Although state-of-the-art rendering methods and graphics hardware can produce realistic images at interactive rates, they do not approach the quality and realism of holographic recordings.

Autostereoscopic displays allow for glass-free observation of computer-generated scenes. These displays can present several perspective views at one time, thus supporting multiple users simultaneously. Resolution and rendering speed, however, decrease with the number of views generated. Holographic images, in contrast, can provide all depth cues—perspective, binocular disparity, motion parallax, convergence, and accommodation—and theoretically can be viewed simultaneously from an unlimited number of positions.

PRINTING AND RENDERING

The two main types of computer-generated holograms, *digital holography* and *electroholography*, have become the focus of several groups researching applications for them.

Digital holography¹ uses holographic printers to sequentially expose small fractions of the photo-

Holography and Autostereoscopy

Both of the two basic optical hologram types—*transmission* and *reflection*—are reconstructed by illuminating them with monochromatic light, as described in Figure A. This approach requires that the light hit the emulsion at the same angle as the reference laser beam used to record the hologram.

Transmission and reflection holograms

To view a transmission hologram, the light source and observer must be on *opposite* sides of the holographic plate. The light is transmitted through the plate before it reaches the observer's eyes. Those portions of the emulsion not recorded or illuminated remain transparent.

To view a reflection hologram, the light source and observer must be on the *same* side of the holographic plate. The light reflects from the plate toward the observer's eyes. As with transmission holograms, unrecorded or nonilluminated portions of the emulsion remain transparent.

These two hologram types have generated a spectrum of variations. Although some holograms can be reconstructed only with laser light, others can be viewed only under white light. Rainbow

holograms, one of the most popular white-light transmission hologram types, diffract each wavelength of the light through a different angle. This lets viewers observe the recorded scene from different horizontal viewing positions but also makes the scene appear in different colors when observed from different points.

In contrast to rainbow holograms, white-light reflection holograms can provide full parallax and display the recorded scene in a consistent—but in most cases monochrome—color for different viewing positions.

It is possible to produce both transmission and reflection variations of color white-light holograms. Usually, this process requires recording the same content on several emulsion layers while exposing each layer to laser light with a different wavelength. When reconstructed, each object wave from each layer contributes its individual wavelength, merging together into a multicolor image.

Parallax displays

These CRT or LCD display screens are overlaid with an array of light-directing or light-blocking elements. Using these elements, the display directs emitted light to both eyes differently—allowing each eye to see individual portions of the displayed image. The observer's visual system interprets corresponding light rays as being emitted by the same spatial point. Dividing the screen space into left and right image portions allows for a glass-free separation of stereo pairs into two or more viewing zones.

Some displays control the parallax array mechanically or electronically, depending on the observer's viewpoint, to direct the viewing zones more precisely toward the eyes. Others generate many dense viewing zones—each showing a slightly different perspective of the rendered scene at the same time. Such displays support multiple users simultaneously but do not yet allow for high frame rates.

One parallax display type includes a barrier that applies an array of light-blocking elements—for example, a light-blocking film or liquid crystal barrier—in front of a screen. The light-blocking elements cover portions of the screen for one eye that are visible from the other eye.

Another parallax example, a lenticular-sheet display, utilizes the refraction of a lens array—consisting of small, cylindrical, prismlike or spherical lenses—to direct the light into the different viewing zones. Images generated with lenticular-sheet displays appear brighter than those displayed on barrier displays. Although prisms and cylinders provide a horizontal parallax only, spherical lenses support a full parallax.

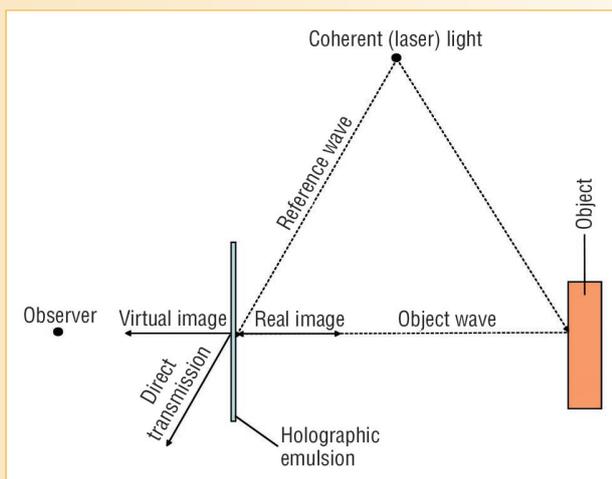


Figure A. Optical holographic recording and reconstruction. A laser beam can be split into two identical beams. One beam, the reference wave, illuminates the holographic emulsion directly while the other beam illuminates the object to be recorded. If the emulsion is illuminated with a copy of the reference wave, it interacts with the recorded interference fringes and reconstructs the object wave, which is visible to the observer.

metric emulsion with a computer-generated image. This process results in conventional holograms that display computer-generated content. This technique also can be used to construct large-scale, tiled holograms.² Although digital holograms can be multiplexed to display scenes in motion, they remain noninteractive.

Electroholography facilitates the computer-based generation and display of holograms in real time.^{3,4} Holographic fringes can be computed by either

- rendering multiple perspective images, then combining them into a stereogram;⁵ or
- simulating the optical interference and calculating the interference pattern.⁶

Once computed, the system dynamically visualizes the fringes with a holographic display. Since creating a hologram requires processing, transmitting, and storing a massive amount of data, today's computer technology still sets electroholography's limits. To

overcome some of these performance issues, researchers have developed advanced reduction and compression methods that create truly interactive electroholograms. Unfortunately, most of these holograms are small, low resolution, and monochrome. However, recent advances in consumer graphics hardware may reveal potential acceleration possibilities that can overcome these limitations.⁷

PARTIALLY RECONSTRUCTING OBJECT WAVES

Combining optical holograms with 2D or 3D graphical elements can provide an acceptable trade-off between quality and interactivity. While the holographic content provides high-quality but static content—as described in the “Holography and Autostereoscopy” sidebar—the combined technology can generate, insert, modify, and animate additional graphical information at interactive rates.

Technically, researchers can use optical combiners such as mirror beam splitters or semitransparent screens to visually overlay the output rendered on a screen over a holographic plate. However, the hologram’s reconstructed light will interfere with the overlaid light of the rendered graphics, making an effective combination impossible.

To solve this problem, researchers can reconstruct the object wave only partially, leaving gaps at those places where they have inserted graphical elements. Doing this requires a point light source capable of selectively emitting light in different directions to create an incomplete reference wave. Conventional video projectors provide such light sources and are well suited to viewing white-light reflection or transmission holograms because today’s high-intensity discharge lamps can produce a very bright light.

If we use autostereoscopic displays, such as parallax displays, to render 3D graphics registered to the hologram, both holographic and graphical content appear three-dimensional within the same space. This effect can also be achieved by using stereoscopic displays with special glasses that separate the stereo images.

Both reflection holograms, which lack an opaque backing layer, and transmission holograms remain transparent if not illuminated. Thus, they can serve as optical combiners themselves—leading to very compact displays. The illumination and rendering techniques work the same for both hologram types.

Figure 1 shows how a transmission hologram can be combined effectively with a flat-panel lenticular-lens sheet display—a variation of a parallax display that utilizes the refraction of a lens array to direct light into the different viewing zones.

Placing a transmission hologram in front of a mir-

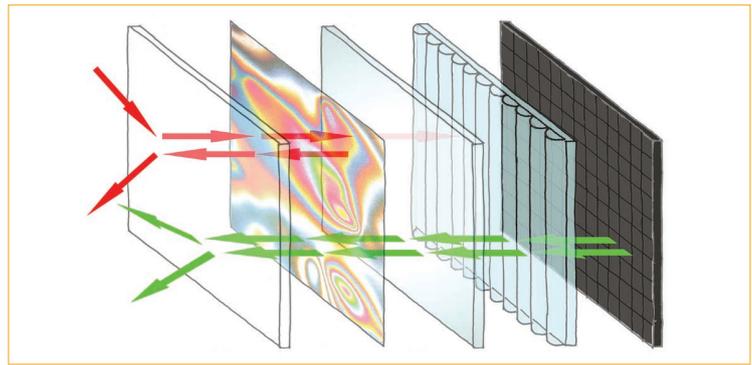


Figure 1. Optical functioning. The explosion model of the optical layers’ stacked structure shows, from left to right: glass protection, holographic emulsion, mirror beam splitter for transmission holograms only, lenticular-lens sheet, and LCD array. Reflected light rays (red arrows) reconstruct the object wave on their return through the emulsion. Stereoscopic images (green arrows) pass through all layers until they merge with the hologram.

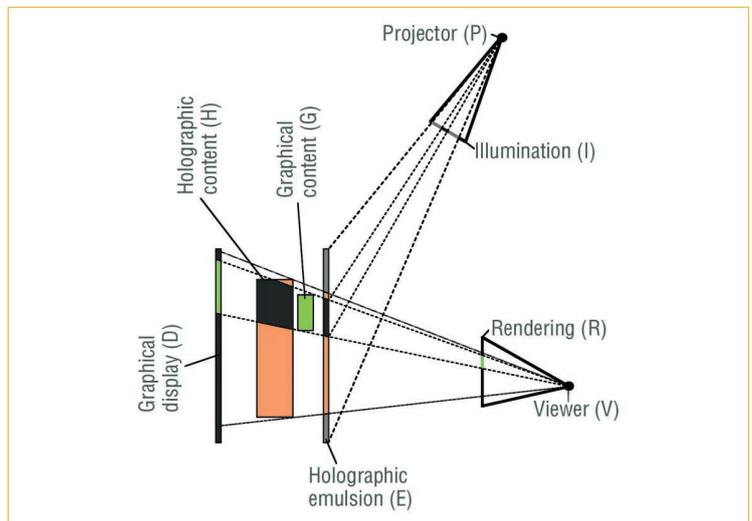


Figure 2. Conceptual sketch of the display constellation. Colored areas on the graphical display and holographic emulsion show visible image areas constitute the hologram (red) and which constitute the graphics (green).

ror beam splitter illuminates it from the front and augments it with graphics from the back. Reflection holograms do not need this beam splitter.

A thin glass plate protects the emulsion from being damaged and keeps it flat to prevent optical distortion. The lenticular-lens sheet directs the light emitted from the LCD array through all layers toward the observer’s eyes. The projected light is transmitted through the first two layers and partially reflected back—either by the beam splitter in combination with a transmission hologram or by a reflection hologram—to reconstruct the recorded content. The screen mostly absorbs the remaining portion of light transmitted through all layers. Figure 2 shows how the system computes the selective illumination on the holographic plate to reconstruct the portion of the hologram not occluded by graphics.

Assuming that information about the holographic content’s depth and a description of the graphical content area are available, researchers can use conventional graphics hardware for ren-

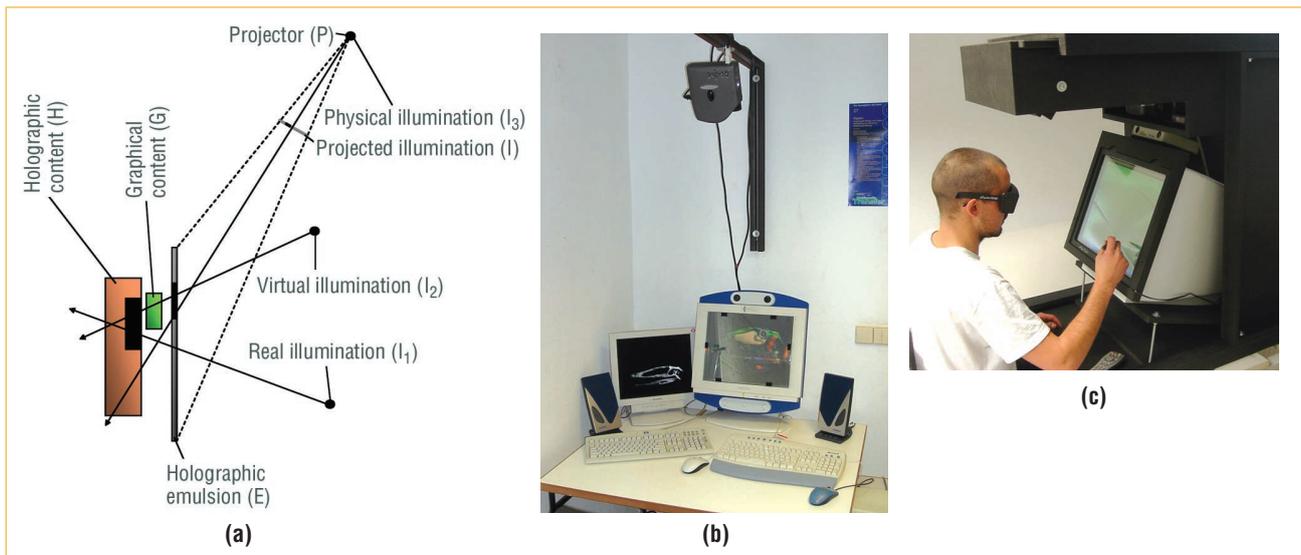


Figure 3. Light-interaction and proof-of-concept prototypes. (a) To simulate virtual shading and shadow effects on the holographic content, the recorded and physical illumination effects must first be neutralized. (b) Autostereoscopic prototype with parallax display and head-finder. (c) Active stereoscopic prototype with CRT monitor, infrared tracking, and touch screen.

dering and illumination. Fortunately, the depth information required to approximate the holographic content's surface can be very coarse. For this example, it consists of a mesh with 5,000 triangles that totals 44.4 Kbytes. This image does not come close to approaching the quality of its optical counterpart but can be rendered easily in real time.

The system geometrically aligns this content during an offline registered step. If the recorded hologram includes optical markers with the actual content, cameras can automatically perform this registration. In addition, the researchers must perform an offline calibration to determine the video projector's extrinsic and intrinsic parameters with respect to the holographic emulsion. If it is mechanically possible to mount the holographic emulsion close to the graphical display, these two planes can be considered identical for the rendering algorithm.

To partially reconstruct the object wave, the system first creates an intermediate texture image by rendering the holographic content from the viewer into the graphics card's depth buffer and filling the card's frame buffer entirely with the lighting source's predefined color values. In addition, the system renders the graphical scene description into depth and stencil buffers, then it clears the frame buffer's stenciled areas and copies the result into the memory block allocated for the intermediate texture image.

If the graphics card provides a render-to-texture option, the read-back operation from the frame buffer into texture memory can be bypassed. The system renders the final illumination image from the projector by drawing the holographic emulsion into the frame buffer and texturing its geometry with the intermediate texture image. The projector then beams the illumination image onto the holographic emulsion.

Next, the system generates a rendering image from the viewer over the off-axis of the graphical display by rendering the content's depth information into

the depth buffer and the graphical content's scene description into the depth and frame buffers. The graphical display then shows the rendering image.

LIGHT INTERACTION

The reconstructed object wave's amplitude is proportional to the reference wave's intensity. Besides using an incomplete reference wave for reconstructing a fraction of the hologram, intensity variations of the projected light permit local modification of the recorded object wave's amplitude.

Practically, this means that to create the illumination image, the system uses shading and shadowing techniques to render the holographic content instead of rendering it with a uniform intensity. To do this, the shading effects caused by the real light sources used for illumination during hologram recording, as well as the physical lighting effects caused by the video projector on the holographic plate, must both be neutralized. Next, the influence of a synthetic illumination must be simulated. This can also be done with conventional graphics hardware, as Figure 3a shows. Three intensity images must be rendered.

For the first image, the system renders the holographic content's depth information from the desired relationship between the viewer and the emulsion, using a white diffuse material factor and graphical light sources that generate approximately the same shading and shadow effects as the real light sources used during the holographic recording process. This results in an intermediate texture. The system generates the first image by rendering the holographic emulsion from the perspective of the video projector and texturing it with the intermediate texture. This image simulates the intensity of the recorded object wave.

The system repeats the process to create the second image, this time using graphical light sources to shade the holographic content under the new, virtual-lighting situation. The ratio of the second image to

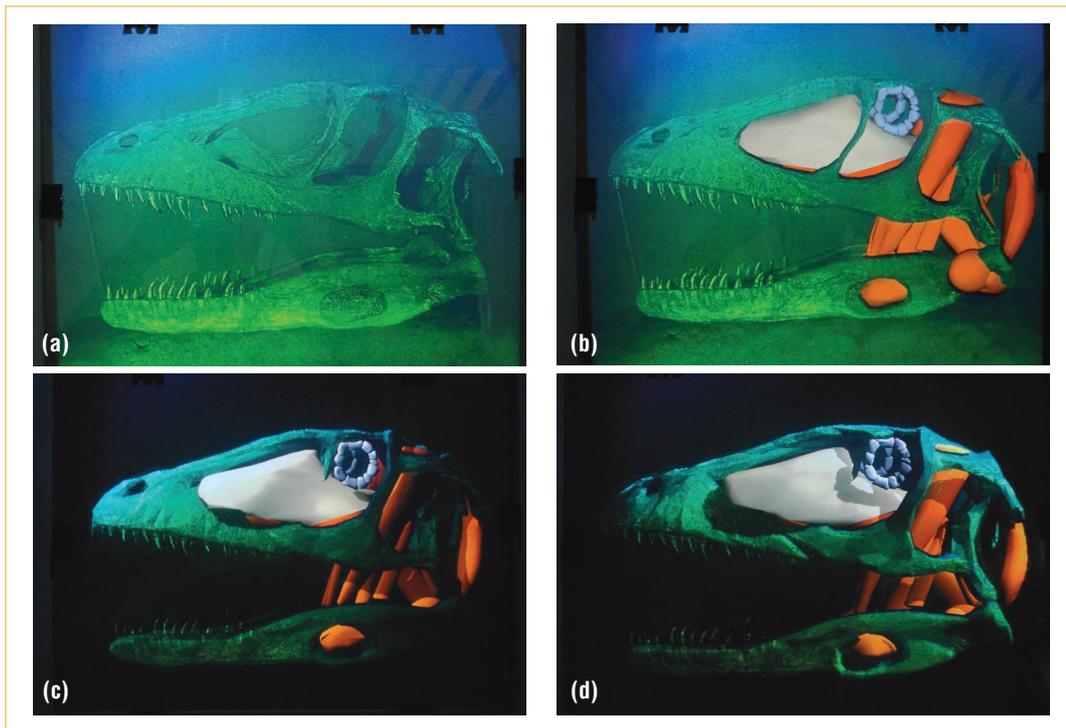


Figure 4. Rainbow hologram of a dinosaur skull combined with 3D graphical elements and synthetic shading effects. (a) The holographic plate illuminated with a uniform light. (b) The plate illuminated only at the portions not occluded by graphical elements such as muscles and other soft tissue. (c) The virtual light source located at the top-left corner, in front of the display. (d) The virtual light source placed at the top-right corner, in front of the display.

the first image represents the required intensity of the reference wave for the holographic plate's emulsion.

For the third image, the system renders the geometry of the holographic emulsion from the projector with a white diffuse material factor and a virtual point light source located at the projector's position. This intensity image represents the geometric relationship between the holographic plate and the video projector as a physical point light source. This third image contains form factor components, such as the square-distance attenuation and angular correlation of the projected light onto the holographic plate, and it neutralizes the physical effects of the projector itself.

The final illumination image can be computed in real time by dividing the second image by the first image and then by the third image via pixel shades. The projection of the resulting image onto the holographic emulsion will neutralize the physical and recorded illumination effects as much as possible and create new shadings and shadows based on the virtual illumination. Again, the system must stencil out the appearance of the graphical content in the final image before displaying it.

During all illumination and rendering steps, the system uses hardware-accelerated shadow-mapping techniques to simulate real and virtual shadow effects on the holographic content's depth information and the graphical scene description. Finally, the system can cast synthetic shadows correctly from all holographic and graphical elements onto all other elements.

This technique's capabilities are clearly limited. It produces acceptable results if the recorded scenery has been illuminated well while making the hologram. However, it cannot neutralize recorded

shadows and extreme shading differences. Further, it cannot cancel out recorded color, reflections, and higher-order optical effects.

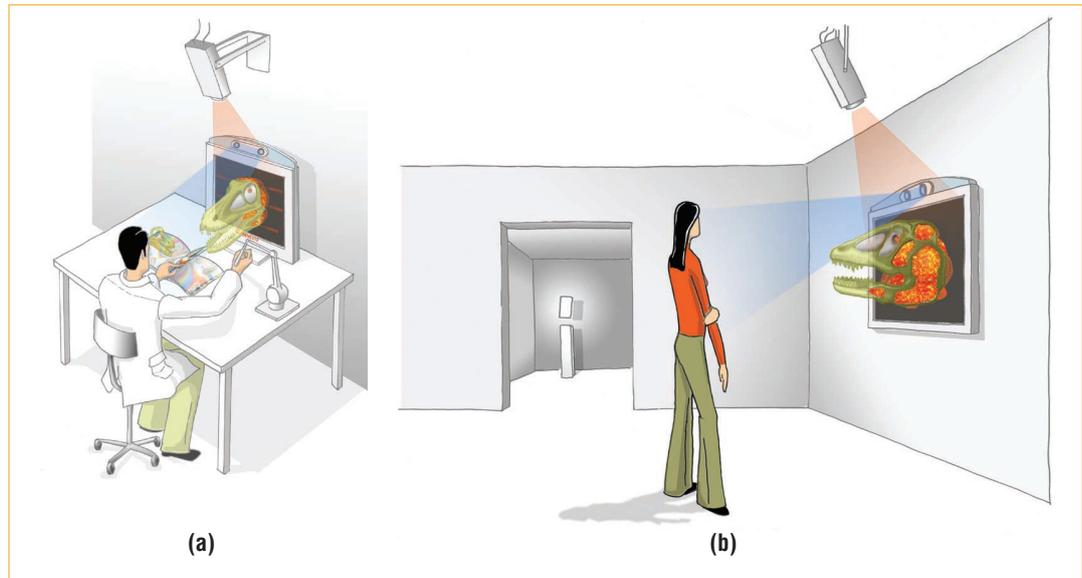
PROVING THE CONCEPT

The autostereoscopic prototype in Figure 3b and the stereoscopic desktop prototype in Figure 3c were built to validate the proposed techniques. These prototypes consist entirely of off-the-shelf components, including either a lenticular-lens-sheet display with integrated head-finder for wireless user tracking or a conventional CRT screen with active stereo glasses, wireless infrared tracking, and a touch screen for interaction. Both prototypes use digital light projectors. A single PC with a dual-output graphics card renders the graphical content on the screen and illuminates the holographic plate on the video projector.

In both cases, the screen additionally holds the front layers—glass protection, holographic emulsion, and optional mirror beam splitter. The remaining two layers shown in Figure 1—the lenticular-lens sheet and LCD array—already form part of the autostereoscopic display. The display shown in Figure 3c does not need them because the system uses shutter glasses to separate the images for the left and right eye.

A rainbow transmission hologram and a reflection hologram of a dinosaur skull were recorded with a 527.5-nm green laser. Figure 4a shows a photograph of the entire reconstructed hologram, illuminated with a projected uniform light. The system generates illumination and stereoscopic images so that the graphical and holographic content can

Figure 5. Display variations. (a) Desktop display that can be used in a light-box fashion, with a special input device allowing interaction and providing haptic feedback of holographic and graphical content. (b) Wall-mounted display in a museum environment, with a ceiling-mounted video projector replacing conventional spot-lights. An integrated head-finder alerts the display when observers stand in front of it.



be merged within the same space, as Figure 4b shows. Projecting an intensity image that contains new shading and shadow effects instead of a uniform illumination lets the system neutralize most of the diffuse shading recorded in the hologram and produced by the projector. The system can then consistently illuminate the holographic and graphical content, creating matching shading and shadow effects, under novel lighting.

Figures 4c and 4d show the synthetic shading effects caused by a virtual light source. In addition, the image shows virtual shadows cast correctly between hologram and graphical elements. Note that none of the photographs have been retouched. When capturing these images, the system rendered the graphics monoscopically. Although Figure 4 shows the results with a monochrome transmission hologram, reflection and color holograms can achieve the same effects.

Given that hardware-accelerated consumer graphics cards support all these rendering techniques, including shadow mapping and shading, using this method can easily achieve interactive frame rates.

INDUSTRY, SCIENCE, AND EDUCATION

Optical holograms can store a massive amount of information on a thin holographic emulsion. This technology can record and reconstruct a 3D scene with almost no loss in quality. For example, the monochrome hologram shown in Figure 4 comprises a theoretical information content of approximately 96.42 Gbytes. Higher-quality holograms can store a fringe pattern that exceeds one terabyte. Reaching the same quality and performance with computer graphics would require rendering the entire data set in real time on a 1.08 terapixel display—equivalent to 9 million points per square millimeter on a 40 × 30 cm panel. For a color hologram of the same size, the information content mul-

tiplies. Moore's law—which asserts that computing power doubles every 18 months—must be applied many times for graphical or electroholographic rendering techniques and displays to reach this quality at interactive frame rates.

A combination of interactive computer graphics and high-quality optical holograms represents an alternative that can be realized today with off-the-shelf consumer hardware. Several commercial online services already offer uncomplicated and inexpensive ways to create color holograms from a set of images or video clips. With this technology, users can create holograms with almost any content—even outdoor scenes.

Archaeologists, for example, already use optical holograms to archive and investigate ancient artifacts.^{8,9} Scientists can use hologram copies to perform their research without having access to the original artifacts or settling for inaccurate replicas. They can combine these holograms with interactive computer graphics to integrate real-time simulation data or perform experiments that require direct user interaction, such as packing reconstructed soft tissue into a fossilized dinosaur skull hologram. In addition, specialized interaction devices can simulate haptic feedback of holographic and graphical content while scientists are performing these interactive tasks. An entire collection of artifacts will fit into a single album of holographic recordings, while a light-box-like display such as that used for viewing x-rays can be used for visualization and interaction, as Figure 5a shows.

This approach has the potential for wide application in other industries. In the automotive industry, for example, complex computer models of cars and components often lack realism or interactivity. Instead of attempting to achieve high visual quality and interactive frame rates for the entire model, designers could decompose the model into sets of

interactive and static elements. The system could record physical counterparts of static elements in a hologram with maximum realism and release computational resources to render the interactive elements with a higher quality and increased frame rate. Multiplexing the holographic content also lets users observe and interact with the entire model from multiple perspectives.

Augmenting optical holograms in museums with animated multimedia content lets exhibitors communicate information about the artifact with more excitement and effectiveness than text labels offer. Such displays can also respond to user interaction. Because wall-mounted variations like the one in Figure 5b require little space, museums can display a larger number of artifacts.

Clearly, holograms or other replicas cannot substitute for original artifacts because viewing those originals is the main reason patrons visit a museum. If, however, a unique artifact is unavailable or too fragile to be displayed, a hologram offers an enticing alternative by showing the artifact as a high-quality, 3D image that, combined with computer graphics, lets users experience it interactively.

We can use this concept to develop a palette of display variations. For example, with only minor changes to the presented rendering techniques, arbitrarily curved holograms such as the cylindrical shapes used for multiplex holograms can be supported instead of only simple planar plates. Even without graphical augmentations, projector-based illumination alone has several potential applications. In combination with optical or digital holograms, it can be used to create visual effects. Certain portions of a hologram, for example, can be made temporarily invisible while others can be highlighted.

Emerging large-scale autostereoscopic displays and existing stereoscopic projection screens will let designers scale up this proposed concept. Not only the display, but also the holograms themselves, can be composed from multiple smaller tiles to reach large dimensions and high resolutions.

The foundations of future 3D displays may have been laid in the late 1940s by Dennis Gabor, who, 20 years later, received the Nobel Prize in physics for his invention of holography. Large interactive electroholographic displays with high resolution and full color would provide the ultimate displays—realizing what today is only possible in science fiction. Technological advances will pave the road toward making this fantasy a reality. Intermediate solutions, such as merging optical

holograms with 3D graphical elements, represent the first steps along this path. ■

Acknowledgments

Thanks to Tim Frieb and Holowood Holografiezentrum Bamberg for their technical support and to Matthias Hanzlik for the illustrations in Figure 5.

References

1. M. Yamaguchi, N. Ohya, and T. Honda, "Holographic 3-D Printer," *Proc. SPIE*, vol. 1212, Springer, 1990, p. 84.
2. M.A. Klug, "Scalable Digital Holographic Displays," *Proc. Image Processing, Image Quality, Image Capture Systems Conf. (IS&T PICS 2001)*, Soc. for Imaging Science and Technology, 2001, pp. 26-32.
3. J.S. Kollin, S.A. Benton, and M.L. Jepsen, "Real-Time Display of 3-D Computed Holograms by Scanning the Image of an Acousto-Optic Modulator," *Proc. SPIE*, vol. 1136, *Holographic Optics II: Principles and Applications*, Springer, 1989, pp. 178-185.
4. M. Lucente, "Interactive Three-Dimensional Holographic Displays: Seeing the Future in Depth," *Siggraph Computer Graphics*, special issue on Current, New, and Emerging Display Systems, ACM Press, 1997, pp. 63-67.
5. M. Lucente and A. Tinsley, "Rendering Interactive Images," *Proc. Siggraph 95*, ACM Press, 1995, pp. 387-394.
6. M. Lucente, "Interactive Computation of Holograms Using a Look-Up Table," *J. Electronic Imaging*, vol. 2, no. 1, 1993, pp. 28-34.
7. C. Petz and M. Magnor, "Fast Hologram Synthesis for 3D Geometry Models Using Graphics Hardware," *Proc. SPIE 03, Practical Holography XVII and Holographic Materials IX*, vol. 5005, Springer, 2003, pp. 266-275.
8. F. Dreesen and G. von Bally, "Color Holography in a Single Layer for Documentation and Analysis of Cultural Heritage," *Optics within Life Sciences (OWLS IV)*, Springer, 1997, pp. 79-82.
9. F. Dreesen, H. Deleré, and G. von Bally, "High-Resolution Color Holography for Archaeological and Medical Applications," *Optics within Life Sciences (OWLS V)*, Springer, 2000, pp. 349-352.

Oliver Bimber is a junior professor for augmented reality at Bauhaus University, Weimar, Germany. His research interests include display technologies, computer graphics, and human-computer interaction. Bimber received a PhD in engineering from Darmstadt University of Technology, Germany. Contact him at bimber@ieee.org.

Feature Article

Superimposing Pictorial Artwork with Projected Imagery

Oliver Bimber, Franz Coriand, Alexander Kleppe,
Erich Bruns, Stefanie Zollmann, and Tobias Langlotz
Bauhaus University Weimar

We present a novel approach for using pictorial artwork as information displays and show how to combine almost any kind of computer-generated visual information directly with the painted content.

Working high above the floor of the Sistine Chapel in the Vatican of Rome, between 1509 and 1512 Michelangelo Buonarroti painted some of the finest pictorial images of all time. On the ceiling of the papal chapel, he created a masterpiece fresco that includes nine scenes from the book of Genesis. Among them is the famous *Creation of Adam* scene—showing God touching Adam's hand. In 1510, an initial study led Michelangelo to draw the Adam figure as a sanguine on a piece of paper. Today, this early drawing is displayed at the London British Museum.

Around 1518 Jacopo Pontormo painted yet another scene from the book of Genesis, called *Joseph and Jacob in Egypt*. It decorated the bedchamber of Pier Francesco Borgherini for many years and is now being displayed at the London National Gallery. This oil painting made headlines after art historians discovered incredible underdrawings beneath the top paint layer. The underdrawings show that the artist had decided to flip the background design after starting work and simply overpainted his initial approach.

Such underdrawings have been found in many Renaissance and Baroque paintings. In 1634, for example, Rembrandt van Rijn painted a self-portrait that was later retouched by one of his students to feature a Russian nobleman. The original portrait was hidden under layers of paint

for more than 300 years, until it was uncovered recently by art restorers. It sold for nearly seven million British pounds.

Pictorial artwork, such as these examples, can tell interesting stories. The capabilities of museums to communicate this and other information, however, are clearly limited. Text legends and audio guides can mediate facts, but offer little potential for presenting visual content such as embedded illustrations, pictures, animations, and interactive elements.

Also because of the difficult economic situation, edutainment is becoming an important factor for museums. By applying new media technologies—such as computer graphics, virtual reality, and augmented reality—exhibit-oriented information might be communicated more effectively, but certainly in a more exciting way.

In this article we describe a novel technological approach, a mathematical model, a real-time rendering algorithm, and examples of presentation techniques for integrating almost any kind of visual information directly into pictorial artwork. Our system displays such information while keeping the observers' attention on the original artifact and doesn't require additional screens.

Technical approach

A seamless and space-efficient way for integrating visual information directly into pictorial artwork is to use the artwork as an information display (see Figure 1). The display can serve as a diffuse projection screen, and we can apply conventional video projectors to show computer graphics together with the painted content. To perceive the projected imagery in the correct colors and intensities, however, requires that the influence of the underlying physical color pigments is neutralized. In most situations, this isn't possible if untreated images are simply projected directly onto arbitrarily colored surfaces. The problem is that the projected light interacts with the color pigments on the canvas and is partially absorbed if the pigment's color isn't fully white.

To solve this problem, we use a new film material that has two properties: first, it's completely transparent and second, it diffuses a fraction of the light projected onto it. The film consists of an even deposition of fine particles on both sides of a polyester base with no visible artifacts. It was used for creating special effects in Hollywood movies such as *Minority Report* and *Paycheck* and sells for \$350 per square foot. Initial



Figure 1. Using pictorial artwork as information displays: (a) concept sketch and (b) first prototype.

measurements have revealed that on average 20 percent (± 1 percent) of the light that strikes the film is diffused while the remaining fraction is transmitted toward the canvas. This 0.1-mm thin, transparent film can be seamlessly overlaid on the canvas (with or without direct contact) by integrating it into the frame that holds the artwork. We use BenQ 7765PA 1,100 ANSI lumens extended graphics array (XGA) digital light projectors to display images on film and canvas.

Mathematical model

In this section we describe how light interacts with a textured surface through the film. If a light beam with incident radiance L is projected onto the transparent film material located on top of the original artwork, a portion d of L is directly diffused from the film while the remaining portion t of L is transmitted through the film. The transmitted light tL interacts with the underlying pigment's diffuse reflectance M on the canvas, and a color-blended light fraction tLM is diffused. The portion $tLMt$ is then transmitted through the film, while the remaining part $tLMd$ is reflected back toward the canvas, where it is color blended and diffused from the same pigment again. This ping-pong effect between film material and canvas is repeated infinitely while for every pass a continuously decreasing amount of light is transmitted through the film that contributes to the resulting radiance R . Mathematically, we can express this as an infinite geometric series that converges

toward a finite value. The same is true for the environment light with incident radiance E emitted from uncontrollable light sources. Because these light sources also illuminate the canvas and the film material, we must also consider the environment light's contribution to R .

Figure 2a (next page) shows this process as a sequence diagram. Note that in contrast to this conceptual illustration, normally no physical gap exists between the film material and canvas, and the light interaction occurs at the same spot.

If all parameters (L , E , M , t , and d) are known, we can compute the resulting radiance R that's visible to an observer in front of the canvas:

$$R = (L + E)d + (L + E)t^2M \sum_{i=0}^{\infty} (Md)^i$$

$$= (L + E) \left(d + \frac{t^2M}{1 - Md} \right) \quad (1)$$

Now that we know R , which we expect to see, we need to solve Equation 1 for L :

$$L = \frac{R}{\left(d + \frac{t^2M}{1 - Md} \right)} - E \quad (2)$$

Equation 2 allows computing the incident radiance L that needs to be projected onto the film and the canvas to create the known result R . The radiant intensity I of the projector to create L is related to a discretized pixel value and is given by

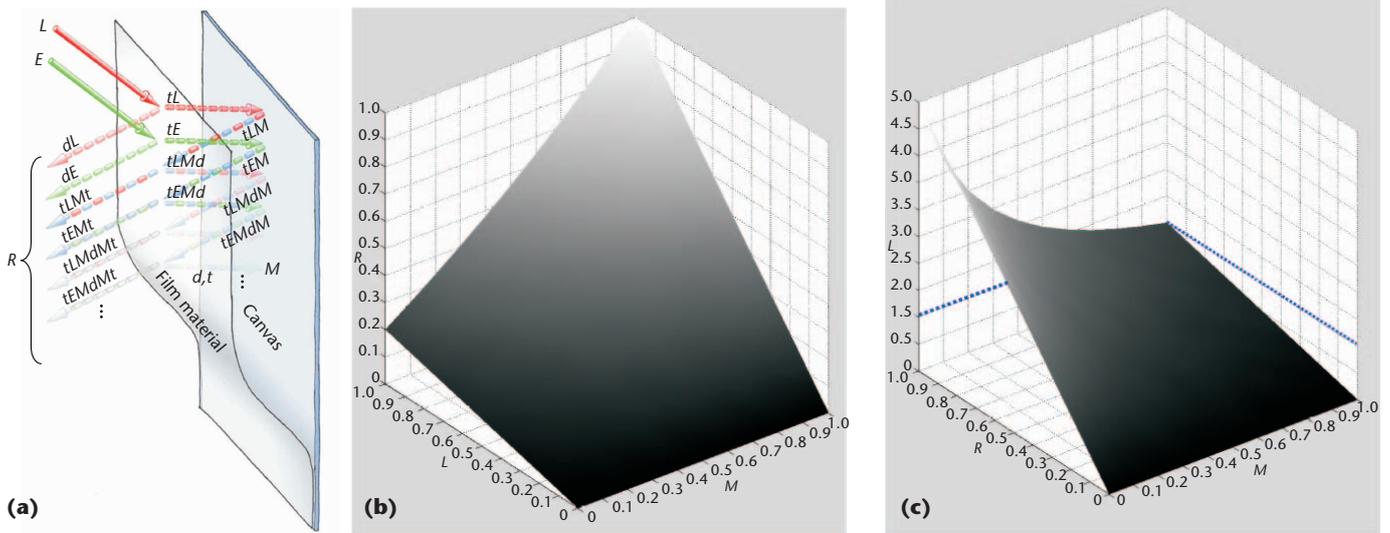


Figure 2. Properties of the film material. (a) Interaction of projected light and environment light with the canvas and the film material (sequence diagram). Visualization of (b) Equation 1 and (c) Equation 2 over R , L , and M (without E , $t = 80$ percent and $d = 20$ percent).

$$I = L \frac{r^2}{\cos \alpha} s \quad (3)$$

where $r^2/\cos \alpha$ are the form factor components: square distance attenuation and angular correlation of the projected light onto the canvas. The additional factor s allows scaling the intensity to avoid clipping and to consider the simultaneous contributions of multiple projectors.

Our approach has clear limitations, which are illustrated in Figures 2b and 2c. Not all radiances R can be produced under every condition. If M is dark, most of L and E are absorbed. In an extreme case, the corresponding pigment is black ($M = 0$). In this case the right term of Equation 1 is canceled out. The remaining left term—which depends on the diffusion factor d of the film material—sets the boundaries of the final result that can be produced. The intersection of the surface with the RL -plane in Figure 2b illustrates these limitations. Consequently, in the worst case of our example, only 20 percent of R can be generated. This situation is also reflected in Figure 2c as the intersection of the surface with the LR -plane. Here we want to assume that $sr^2/\cos \alpha = 1$, which results in $L = I$. For a single video projector, the projected radiance L and consequently the radiant intensity I cannot exceed the normalized intensity value of 1 (dotted line). But for creating most of the resulting radiance values, L and I must be larger. This situation worsens for $r^2/\cos \alpha > 1$ and for $E \rightarrow 1$ or $M \rightarrow 0$.

However, the contributions of multiple (n) projectors allow displacing this boundary with

$$L = \sum_{i=1}^n L_i, \quad I_i = L_i \frac{r^2}{\cos \alpha_i} s_i \quad (4)$$

If all projectors provide linear transfer functions (for example, after gamma correction) and identical brightness, $s_i = 1/n$ balances the load among them equally. However, s_i might be decreased further to avoid clipping and to consider differently aged bulbs.

For Figures 2b and 2c, we don't consider the environment light E and set it to zero. Additional environment light would simply shift the surface in Figure 2b up on the R -axis, and the surface in Figure 2c down on the L -axis. Note that our mathematical model must be applied to all color channels (such as red, green, and blue for projected graphical images) separately.

Real-time color correction

We implemented Equations 2 through 4 as a pixel shader to support real-time color correction. We used Nvidia's Cg framework for fragment processing on an MSI GeForce FX 5600 TD graphics board. Figure 3 illustrates the rendering process based on an example of Michelangelo's *Creation of Adam*. Although we'll use this example to explain the process, it's universal and can be applied with arbitrary background images.

In our example, a copy of the original Adam drawing serves as a background image. Our goal is to overlay it entirely with a registered view on the actual ceiling fresco of the Sistine Chapel.

The first step of the rendering process is to create an input image I_i , which needs to be overlaid.

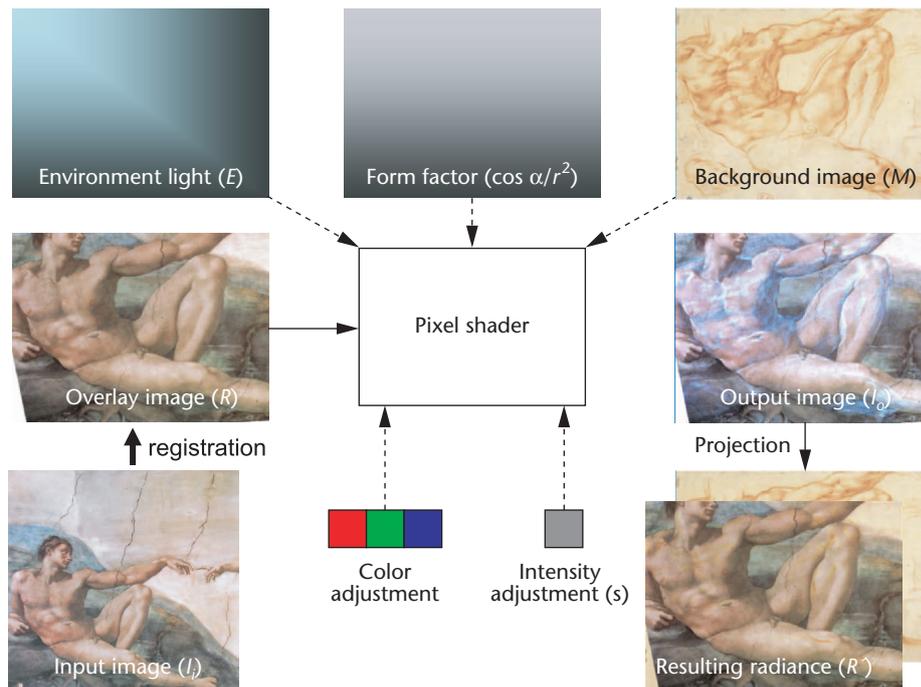


Figure 3. Real-time color-correction process.

This image can be dynamically rendered (as part of a real-time animation or an interactive experience), played back (frames of a prerecorded movie), or static (a photograph of the corresponding ceiling portion—as is the case in our example).

The input image must be registered to the physical drawing. Registration is achieved by texture mapping I_i onto a predistorted image geometry that's precisely aligned with the physical drawing. The amount of distortion depends on the geometric relation between the video projector and the canvas. The distortion can be simple keystone deformations to more complex curvilinear warping effects (for example, if lens distortion of the projector has to be neutralized).

Several automatic approaches apply video cameras and image analysis to align multiple images of tiled projection screens.^{1,2} A structured light registration benefits from controllable features, such as projected grid edges or Gaussian matchpoints that can be easily detected in the camera views with a subpixel precision.

In our case, however, we have to register a digital representation of the artistic content against its physical representation on the canvas, rather than registering one projected structured light image against another. To detect nonstructured artistic features (such as fine lines) in the artwork and register them automatically against the corresponding features in the digital content represents an important task of computer vision—especially

if projected pixels and physical pigments must align precisely on the canvas. We're critical about the feasibility and precision of an automatic method for our problem and have decided to solve it with a manual registration process that benefits from the resolution of the human visual system. Because the following steps have to be performed only once, we believe that they represent an acceptable solution.

Our manual registration process lets users interactively identify 2D correspondences between artistic features in the background image M within the image space—that is, an image of the drawing displayed on a control screen—and the physical drawing on the wall. This is done using a 2D input device, such as a conventional mouse whose pointer is visible on the control screen and as projection on the canvas.

We use dual output graphics card and an additional video signal splitter to drive the control screen and one or two projectors. The result of this feature matching is a set of 2D vertex fiducials with their corresponding texture coordinates within the image space. The fiducials are Delaunay triangulated, and the system uses texture coordinates to map the correct image portions of I onto the image geometry. This results in the overlay image R . Here we should stress again that a precise correspondence between R and M is important to achieve qualitatively good results.

In our experiments, the measurement of 50 to

70 fiducials proved sufficient. In contrast to uniform grid methods normally applied for projector alignment, this general geometric registration allows correlating arbitrary artistic features in the physical drawing with the corresponding pixels of M in the image space.

Thus, it provides an accurate matching that can be regionally improved further if linear interpolation within single-grid triangles fails to be precise enough. The registration results don't change if the projector and background image are fixed. Before we can render R , we must enable the color-correction pixel-shader. Five parameters are passed to the pixel shader to ensure that Equations 2 through 4 can be computed.

The first parameter is the environment light E in the form of intensity texture that has the same size as R . It contains intensity values that represent the uncontrollable lighting situation on the canvas. We can determine the intensity values by measuring the irradiance of the environment light with a light meter for a discrete number of sample spots on the canvas' surface—resulting in the lux (lx) values E' . These values have to be normalized to an intensity space that ranges from 0 to 1 so the shader can process them. To do this, we measure the same spots again, but this time the highest intensity possible (for example, a white image) is projected onto the light meter, which is measured in addition to the environment light. These measurements are equivalent to the total irradiance $T' = L' + E'$, and also carry the unit lux.

Because we know that $L' = T' - E'$ is equivalent to the scaled intensity value $\cos \alpha/r^2$, we can convert the measured radiance of the environment light from lux into the normalized intensity space with $E = E'/(T' - E') \cos \alpha/r^2$. To approximate the intensity values for the entire image area in E , all the measured spots are mapped onto the image space and are Delaunay triangulated. The values for the remaining pixels are linearly interpolated by the graphics card while rendering the Delaunay mesh. Note that E is constant if the environment light doesn't change. For reasons that we describe next, we can assume that $\cos \alpha/r^2$ is constant and equals 1.

The second parameter is the form factor that represents the geometric relation between the video projector as a point light source and the canvas. Because it doesn't change for a fixed relation between projector and canvas, we can precompute it and pass it to the pixel shader in the form of an intensity texture with the same dimensions as E and R .

Like Raskar's Shader Lamps,³ the graphics pipeline can produce this texture—a geometric model of the canvas is rendered with a white diffuse reflectance from the viewpoint of the projector. Attaching a virtual-point light source (also with a white diffuse light component) to the position of the projector and enabling square distance attenuation produces intensities proportional to $\cos \alpha/r^2$. The pixel shader can compute the required reciprocal. Practically (that is, for normal-sized canvases and nonextreme projector configurations), we found that the form factor is constant over all pixels. It's then contained by the intensity adjustment parameter s .

The third parameter is the background image M . It also has the same dimensions as E and R . This image can be generated by, for example, scanning the color values or taking a photograph of the original drawing under uniform illumination.

The fourth and fifth parameters contain color and intensity adjustment values that let users fine-tune the video projector's individual color response and prevent intensity clipping. These parameters also help users adopt color drifts that they can introduce while capturing the background image and allow users to consider the contributions of multiple projectors. Note that gamma correction must be applied in advance. This mostly occurs with projectors with nonlinear transfer functions as well as projectors with linear transfer functions that apply a de-gamma mapping on the video input. Gamma correction is usually supported by the graphics hardware and the video driver, but the pixel shader can also carry it out. We adjust these values manually, but the support of automated methods for color matching multiple projectors⁴ is imaginable.

The output image I_o is the final result of this rendering process and will be displayed by the video projector. If projected geometrically correctly onto the drawing, the result R' will be visible. Both images— R and R' —are mostly identical, except for slight artifacts that appear because of previously discussed limitations. Figure 4 shows the results of our example projected onto a real drawing with a single video projector.

Apparently, the underlying drawing can be made partially or completely invisible to display the graphical overlay in the correct colors on top of it. Figures 4e through 4h show close ups in which diverging body parts (such as belly and knee) are overdrawn and displaced by the projection.

Some intensities and colors that are required

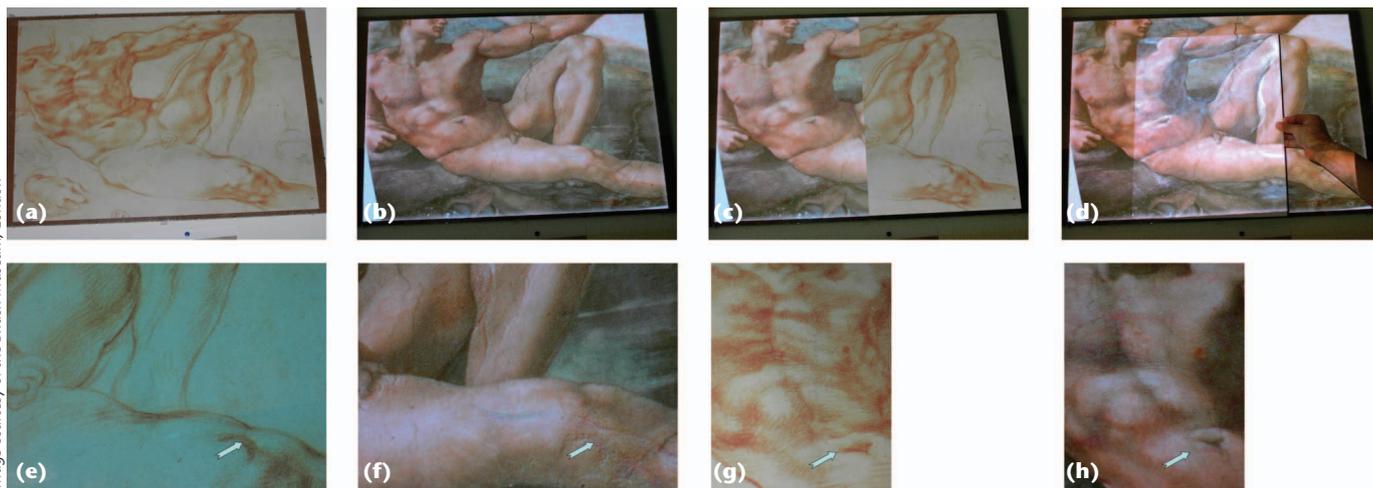


Figure 4. Results of the color-correction process with a single projector on a real drawing. (a) Real drawing (64 × 48 cm) under environment light. (b) Output image emitted onto drawing. (c) Partially augmented drawing. (d) Output image on a white piece of paper. (e–h) Close ups. While the upper body coincides with the drawing and painting, Michelangelo modified the lower body. The arrows indicate the displaced knee and belly sections. They point at the same spot on the drawing.

to neutralize the underlying color pigments can't be achieved by a single video projector. The worst case is to turn a black pigment on the canvas into a white color spot. Figures 2b and 2c illustrate that in such a case the required intensity can easily exceed the boundary of 1 in our normalized intensity space. The pixel shader clips these values to 1, which results in visible artifacts.

The simultaneous contributions of multiple projectors can reduce or even eliminate these effects. Figure 5 shows the extreme case of an input image that has no geometric correspondences to the underlying background image. In addition, both projections together create bright colors (the sky) on top of dark color pigments on the canvas. In Figure 5a, a single projector is used. Intensity values that are too large are clipped and result in visible artifacts. Balancing the intensity load between two projectors reduces these artifacts clearly (see Figure 5b). Figures 6 through 8 (next page) show more results for other cases.

Because of the hardware acceleration of today's graphics cards, we can easily perform the color-correction process in real time. Note that none of the photographs in this article have been retouched. Slight variations in color and brightness are because of different camera responses.

Content creation, authoring, and presentation

The basic color-correction process allows visualizing all sorts of information and effects by ren-

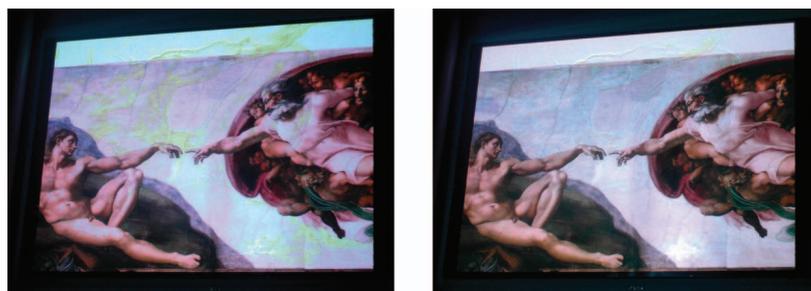


Figure 5. Results of color-correction process with two projectors. (a) Limited intensity capabilities of a single projector result in visible artifacts. (b) The contribution of a second projector reduces these effects.

dering the desired result as input image I_i into the pixel shader. This opens the potential for applying a wide range of established or new presentation tools and techniques. We want to categorize them into six main classes:

- inlay objects (textual information, images, videos, and arrow annotations);
- lens effects (magnification, x-ray, toolglasses, and magic lenses⁵);
- focus effects (blurring, decolorization, hiding, and highlighting);
- 3D effects (walk- or flythrough and object observation);
- modification effects (drawing style modifica-



Figure 6. Sample scenes of the papal chapel's ceiling projected onto the Adam drawing (Figure 4a) during an interactive slide presentation. The underlying drawing is barely visible.



Figure 7. Pontormo's Joseph and Jacob in Egypt (65 × 49 cm). (a) Copy of original painting illuminated under environment light, (b) modification of painting style from oil on wood to watercolor on paper via a 2D artistic filter, (c) re-illumination and lens flare, and (d) registered visualization of underdrawings (infrared recordings are black and white).

Image courtesy of the Vatican Museum, Rome

Image courtesy of the National Gallery, London

Figure 9 illustrates a variety of examples. Embedded information windows (see Figure 9a), such as the semitransparent text panel or the opaque image and video panel are—in contrast to simple physical text labels—dynamic and can be directly linked to corresponding portions of the artwork (via arrow annotations, for example). The magnifying glass in Figure 9b allows zooming into interesting areas to identify brush strokes and hatching styles of the artist. Usually museum visitors are restricted from getting close enough to recognize such details. To draw the observers' attention to a certain image portion, it can be brought into the foreground while the remaining part is eclipsed. In Figure 9c, the focus area is highlighted while the rest is decolorized.

This is technically achieved by projecting the inverse colors of the background image onto *M*. In this case, the colors of the pigments on the canvas are physically canceled out and *M* appears in grayscale. Figure 9d shows a 3D flythrough in the Sistine chapel—building a spatial correlation to surrounding paintings and the environment. This allows relocating the observer virtually to remote places.

Stereoscopic rendering (and optional head-tracking technology) allows an immersive experience. Although the transparent film preserves the polarization of light, the underlying canvas doesn't. Consequently, we can only apply active or anaglyphic stereo solutions. We use a mechanical (half opaque and half transparent) shutter wheel rotating with 120 Hz in front of two video projectors. A light sensor measures the wheel's position and triggers the shutter glasses' infrared synchronization signal. In combination with conventional LCD projectors, this is a simple and—compared to CRT projectors—cost-efficient alternative for active stereo.

Figure 7 illustrates further presentation examples, such as modifications of the painting style via 2D artistic filters, visualization of the underdrawings, and scene modification through re-illumination.

For noninteractive presentations, it's possible to pregenerate the entire content. In this case, we can apply well-established content creation and authoring tools (such as digital imaging, 3D modeling, video editing, and audio-mixing packages). These tools already provide techniques such as animations, image filtering, rollover effects, and so on, as well as professional graphical user interfaces. For interactive presentations, however, the generated content must be man-

tion, re-illumination, and color restoration⁶); and

- audio effects (verbal narrations, music, and sound that are synchronized to the visual effects).

aged by the presentation software linked to an interaction framework.

If no user interaction is required, we apply an embedded Audio Video Interleaved (AVI) player to map video frames onto input images I_i on the fly. This represents a direct interface to our own player framework that comprises rendering and color correction, and allows the user or content creator to benefit from the capabilities of established content creation and authoring packages, such as Adobe Photoshop, Discreet 3ds max, Alias Maya, or Adobe Premiere. Our content interface for 3D interactive and stereoscopic/head-tracked presentations is Nvidia's NVB format, which can be exported from Discreet's 3ds max and displayed by our player.

To see how others have approached incorporating projectors, see the sidebar on the next page, "Previous and Related Work."

Future work

Using pictorial artwork as information displays opens another door to embedded multimedia content in museums and art galleries. Our proposed method is simple, seamless, cost and space efficient, robust, and compatible with off-the-shelf hardware and software. These are all important factors for museum operators, content creators, and museum visitors. The presented techniques let us think of a wide variety of presentation and interaction tools that we didn't explore in this article. Dynamic view management and label placement are only two examples. As for other areas, the development of efficient interaction techniques and devices for such displays will be an interesting challenge.

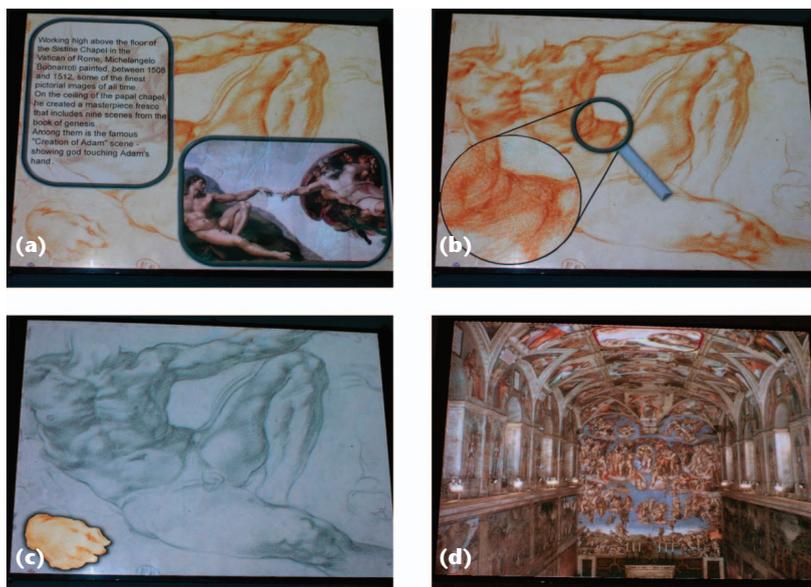
As we previously discussed, our method has limitations that are mainly defined by the capabilities of the applied hardware. For example, the restricted resolution, brightness, contrast, minimum focus, distance, and black level of video projectors are issues that will certainly be improved by next-generation projectors. The XGA resolution and the brightness of 1,100 ANSI lumens of our low-cost projectors were appropriate for small- and medium-sized paintings in a normally lit environment. An upscaling is possible by using more projectors, but downscaling would either result in a loss of effective resolution or in focus problems.

Black, for instance, is a color that can't be projected. Instead the environment light together with the black level of the projectors illuminates areas that need to appear black. We found that



Image courtesy of the Museum het Rembrandthuis, Amsterdam

Figure 8. Rembrandt's self-portrait (48 × 56 cm). (a) Copy of original painting as it looks today (illuminated under environment light). Various cleaning stages to remove the overpainted layers from (b) 1935, (c) 1950, and (d) 1980 are projected onto (a). Only black-and-white photographs of these stages are available. The high black level of the video projectors prevents us from creating a totally black color on the canvas. Extreme regions, such as overlaid hair and a hat can't appear completely black for this reason.



Images courtesy of the British Museum, London, and the Vatican Museum, Rome

Figure 9. Examples of presentation techniques: (a) inlay text and image, (b) magnification, (c) focus through highlighting and decolorization, and (d) 3D flythrough in the Sistine chapel. Note that in images (a–c) the Adam drawing itself isn't projected.

Previous and Related Work

With increasing capabilities and decreasing costs, video projectors have become widespread and established presentation tools. Being able to generate images that are larger than the actual display device virtually anywhere is an interesting feature for many applications that can't be provided by desktop screens. Several research groups have discovered this potential by applying projectors in unconventional ways to develop new and innovative information displays that go beyond simple screen presentations.

The Luminous Room,¹ for instance, describes an early concept for providing graphical display and interaction at each interior architecture space's surface. Co-located two-way optical transducers—called *I/O bulbs*—that consist of projector-camera pairs capture the user interactions and display the corresponding output. With the Everywhere Displays Projector,² Pinhanez has extended this concept technically by allowing a steerable projection using a pan/tilt mirror. Later, Raskar et al.³ demonstrate how context-aware handheld projectors—so-called *iLamps*—can be used as mobile information displays and interaction devices.

Raskar et al. also use multiple projectors for their Shader Lamps⁴ approach to lift the visual properties of neutral diffuse objects that serve as a projection screen. The computed radiance at a point of a nontrivial physical surface is mimicked by changing the bidirectional reflectance distribution function (BRDF) and illuminating the point appropriately with projector pixels. Animating the projected images lets us create the perception of motion without physical displacement of the real object.⁵ This type of spatial augmentation is also possible for large, human-sized environments, as demonstrated by Low et al.⁶

Projector-based illumination has become an effective technique in augmented reality to achieve consistent occlusion^{7,8} and illumination⁹ effects between real artifacts and optically overlaid graphics. Video projectors instead of simple light bulbs are used to illuminate physical objects with arbitrary diffuse reflectance. The per-pixel illumination is controllable and can be synchronized with the rendering of the graphical overlays. It also makes the combination of high-quality optical holograms with interactive graphical elements possible.¹⁰ Using a video projector to produce a controlled reference wave lets us reconstruct a hologram's object wave partially—but not at those por-

tions that are overlaid by integrated graphical elements.

Yoshida et al.¹¹ describe a virtual retouching technique that applies video projectors for reconstructing and enhancing the faded colors of paintings by projecting light directly onto the canvas. An affine correlation between projection and the result captured by a camera is established manually for each pixel. Users can then retouch the original painting interactively via a desktop GUI.

In the context of our own approach, we can divide these methods into four main groups:

- information displays that project arbitrary images onto surfaces with arbitrary reflectance, but don't consider the effect of color blending;¹⁻³
- techniques that project colored images onto surfaces with a neutral white reflectance to avoid color blending;⁴⁻⁶
- methods that project a uniformly colored illumination onto surfaces with arbitrary reflectance and texture;⁷⁻¹⁰ and
- approaches that project colored images onto surfaces under consideration of their reflectance.¹¹

As with Yoshida's work, our approach belongs to the fourth group. Although there are basic conceptual similarities between both attempts, they differ in their general technological realization, aimed applications, as well as in their mathematical model and rendering techniques. We want to know whether a direct projection of colored light onto surfaces with arbitrary reflectance will work effectively under real-world conditions. The limitations of such an approach will be reached quickly in cases where an adequately large portion of the color spectrum is absorbed by the underlying surface and can consequently not be reflected.

References

1. J. Underkoffler, B. Ullmer, and H. Ishii, "Emancipated Pixels: Real-World Graphics in the Luminous Room," *Proc. ACM Siggraph*, 1999, ACM Press, pp. 385-392.

even in our case the human vision system adjusts well to local contrast effects—which makes these areas appear much darker than they actually are. Even with little environment light, however, the high black level of video projectors causes this illusion to fail in extreme situations, such as the one shown in Figure 9. The development of video projectors indicates that we can expect a decrease of the black level and an increase of the contrast ratio in the future.

Light can damage the artwork. Ultraviolet (UV) and infrared (IR) radiation produced by the lamps of video projectors is critical. Commercially available UV/IR blocking filters can be mounted in front of the projectors' lenses to remove most of these unwanted rays while transmitting visible wavelengths.

For the remaining visible light portion, a general rule of thumb in the museum field advises to illuminate valuable and delicate pieces perma-

2. C. Pinhanez, "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces," *Proc. Ubiquitous Computing*, Springer, 2001, pp. 315-331.
3. R. Raskar et al., "iLamps: Geometrically Aware and Self-Configuring Projectors," *Proc. ACM Siggraph*, ACM Press, 2003, pp. 809-818.
4. R. Raskar et al., "Shader Lamps: Animating Real Objects with Image-Based Illumination," *Proc. Eurographics Rendering Workshop*, Eurographics Assoc., 2001, pp. 89-102.
5. R. Raskar, R. Ziegler, and T. Willwacher, "Cartoon Dioramas in Motion," *Proc. Int'l Symp. Nonphoto-realistic Animation and Rendering*, ACM Press, 2002, pp. 7-ff.
6. K. Low et al., "Life-Sized Projector-Based Dioramas," *Proc. Symp. Virtual Reality Software and Technology*, IEEE CS Press, 2001, pp. 93-101.
7. O. Bimber and B. Fröhlich, "Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality*, IEEE CS Press, 2002, pp. 186-195.
8. S. Noda et al., "An Optical See-Through Mixed Reality Display with Real-Time Rangefinder and an Active Pattern Light Source," *Trans. Virtual Reality Soc. of Japan*, Virtual Reality Soc. of Japan, vol. 4, no. 4, 1999, pp. 665-670.
9. O. Bimber et al., "Consistent Illumination within Optical See-Through Augmented Environments," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality*, IEEE CS Press, 2003, pp. 198-207.
10. O. Bimber, "Combining Optical Holograms with Interactive Computer Graphics," *Computer*, vol. 37, no. 1, Jan. 2004, pp. 85-91.
11. T. Yoshida, C. Horii, and K. Sato, "A Virtual Color Reconstruction System for Real Heritage with Light Projection," *Proc. Virtual Systems and Multimedia*, IEEE CS Press, 2003, pp. 158-164.

nently with no more than 100 lx–150 lx. The potential damage caused by light is cumulative (for example, 1 hour with 1,000 lx equals 1,000 hour with 1 lx). The damage also depends on the material and color of the painting and the wavelength (color) of the light. A temporary illumination of a higher light intensity isn't critical.

To give rough indications of the light levels that were reached in our experiments, we measured an ~300 lx maximum for the results shown

in Figure 4 (single projector), an ~400 lx maximum for the results shown in Figure 6 (two projectors), and an ~700 lx maximum for the results shown in Figure 5 (also two projectors).

During a 2–3 minute presentation, this increased lighting is only temporary and such highlight situations usually only appear (if at all) for a short period of time (a few seconds) at varying locations on the canvas. Nevertheless, using the intensity adjustment we've described, the maximum light level can be constrained to be below an upper threshold. However, this might cause visible artifacts depending on the presented content, the painting, and the environment light (as described in Figure 2). Thus, it's important to reach a good balance between total illumination (projected and environment light) over time and creating convincing presentation effects.

We currently consider only the intensity of the environment light E . This is adequate for regular white light sources but will result in artifacts if visible color shading is created on the canvas by the environment illumination. Without modifying the mathematical model or rendering process, the environment light's color can be compensated by determining it with the aid of colorimeters, encoding this information in E , and passing it to the pixel shader.

Using cameras in combination with projected structured light samples might help us automate our registration, color, and intensity adjustment steps. This, however, requires that an acceptable level of precision can be achieved.

The presented concept and techniques are applicable in combination with diffuse pictorial artwork, such as watercolor paintings, pen or ink drawings, sanguine sketches, or matte oil paintings. Extreme light and view-dependent effects, such as non-Lambertian specular reflections; self-shadows; subsurface scattering and interreflections that are created by brush strokes, paint material; or canvas textures can't be handled yet. We'll investigate these ideas in our future research. However, some of these cases will require components—such as head tracking—that might not be effectively integrated into museum environments.

We plan to evaluate our approach in museums and art galleries. We'll have to conduct user studies and surveys on the efficiency and acceptability of this technology in the future. The feedback from visitors, curators, and staff members will lead to improvements and new ideas for the interactive presentation of pictorial artwork. **MM**

Acknowledgments

We thank the Museum het Rembrandthuis, Amsterdam, for providing digital photographs used for the experiment shown in Figure 8, and Matthias Hanzlik for the illustrations in Figures 1 and 2.

References

1. Y. Chen et al., "Automatic Alignment of High Resolution Multi-Projector Displays Using an Uncalibrated Camera," *Proc. IEEE Visualization*, IEEE CS Press, 2000, pp. 125-130.
2. R. Yang et al., "PixelFlex: A Reconfigurable Multi-Projector Display System," *Proc. IEEE Visualization*, IEEE CS Press, 2001, pp. 21-26.
3. R. Raskar et al., "Shader Lamps: Animating Real Objects with Image-Based Illumination," *Proc. Eurographics Rendering Workshop*, Eurographics Assoc., 2001, pp. 89-102.
4. A. Majumder et al., "Achieving Color Uniformity Across Multi-Projector Displays," *Proc. IEEE Visualization*, IEEE CS Press, 2000, pp. 117-124.
5. E. Bier et al., "Toolglasses and Magic Lenses: The See-Through Interface," *Proc. ACM Siggraph*, ACM Press, 1993, pp. 73-80.
6. T. Yoshida, C. Horii, and K. Sato, "A Virtual Color Reconstruction System for Real Heritage with Light Projection," *Proc. Virtual Systems and Multimedia*, IEEE CS Press, 2003, pp. 158-164.



Oliver Bimber is a junior professor for augmented reality at the Bauhaus University Weimar in Germany. His research interests include future display technologies, real-time rendering, optics, and advanced human-computer interfaces. He holds a PhD in computer science from the Technical University of Darmstadt.



Franz Coriand is a graduate student of media systems at the Bauhaus University Weimar. His research interests include augmented reality and virtual reality, as well as software engineering and Unix-based operating systems.



Alexander Kleppe is a graduate student of media systems at the Bauhaus University Weimar. His research interests include augmented reality and computer graphics, as well as input devices and interaction techniques for virtual environments.



Erich Bruns is a graduate student of media systems at the Bauhaus University Weimar. His research interests include computer graphics, augmented reality, and the psychology of perception, as well as the relation between anthropology and computer science.



Stefanie Zollmann is a graduate student of media systems at the Bauhaus University Weimar. Her research interests include virtual reality and augmented reality, as well as input and manipulation techniques for 3D environments.



Tobias Langlotz is a graduate student of media systems at the Bauhaus University Weimar. His research interests include augmented reality, virtual reality, and multiuser game engines.

Readers may contact Oliver Bimber at obimber@computer.org.

For further information on this or any other computing topic, please visit our Digital Library at <http://www.computer.org/publications/dlib>.

COVER FEATURE

Embedded Entertainment with Smart Projectors



Essentially video projectors enhanced with sensors to gain information about the environment, smart projectors do not require artificial canvases and allow correct projection of images onto many arbitrary existing surfaces, such as papered walls or curtained windows.

Oliver
Bimber

Andreas
Emmerling

Thomas
Klemmer

Bauhaus University
Weimar

Television played a central role in shaping the 20th century and remains the primary entertainment medium for most of us today. TV continues to evolve as innovative display technologies change its look and capabilities at an accelerating rate. The popularity of today's flat-panel liquid-crystal and plasma displays shows that emerging trends favor large-screen displays. Simultaneously, falling prices have led to a booming market for home entertainment technology. However, the physical limitations inherent in these technologies place constraints on maximum screen size, display size, refresh rate, and power consumption.

Another display type may soon conquer the entertainment market, however: *Video projectors* have experienced an enormous metamorphosis during the past decade. The cost reductions and performance increases made in these devices compare favorably with those personal computer manufacturers achieved decades earlier. Video projectors also offer a vital advantage over other display technologies. They can generate images much larger than the devices themselves without being constrained by a traditional TV screen's limitations.

This ability comes at a price, however: The artificial canvas requires a space equal to the size of the image we want displayed. A home theater, for example, might require an entire room. In many situations, the temporary or stationary canvases that projector-based multimedia presentations require also harm the ambience of environments such as a living room or historic site.

Smart projectors, however, do not require an artificial canvas. Instead, they allow a correct projection onto many arbitrary existing surfaces, such as papered walls or curtained windows.

SMART PROJECTORS

Essentially video projectors enhanced with sensors to gain information about the environment, *smart projectors* primarily use cameras to sense their environment. However, other information gathering devices such as *tilt sensors* are also available. Completely calibrated and mounted as a single camera projector unit, or realized with separated components, some smart projectors allow dynamic elimination of shadows the user casts,¹ automatic keystone correction on planar screens,² or manually aligned shape-adaptive projection on second-order quadric display surfaces³ such as cylinders, domes, ellipsoids, or paraboloids.

For projection planes, cameras can help to automatically register multiple projector units based on homographic relationships.⁴ In this case, camera feedback also provides the data for intensity blending and color matching⁵ of multiple projector contributions. Combining calibrated stereo cameras with projectors allows direct scanning of an arbitrary display surface's 3D geometry, enabling undistorted projection for a known head-tracked observer position.⁶

All these approaches require the calibration of cameras and projectors to determine their intrinsic position—the focal length, principal point, skew

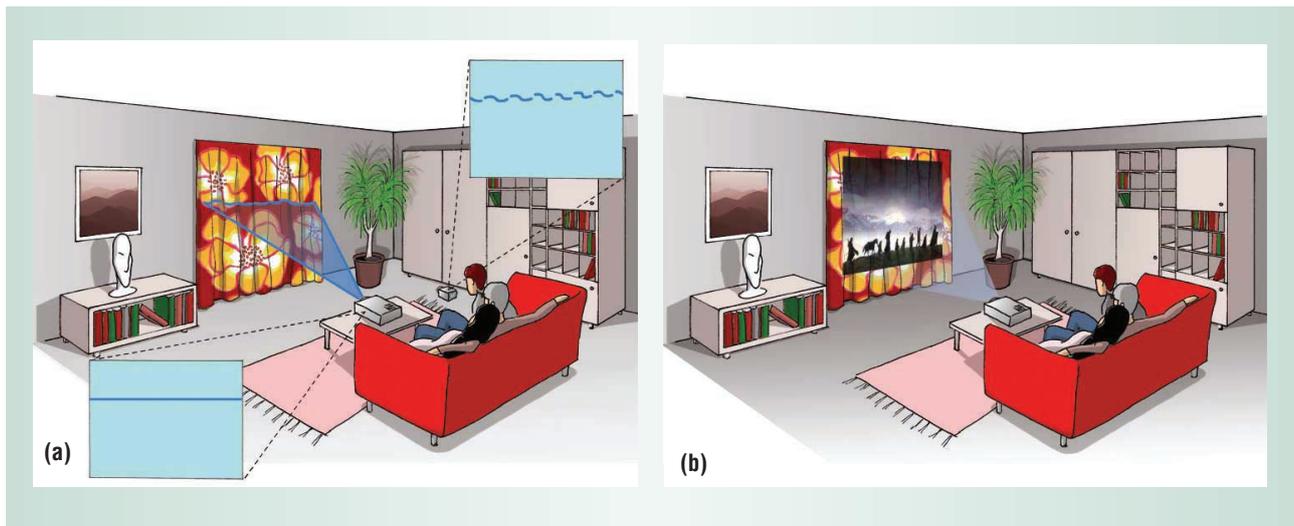


Figure 1. Smart projector concept. The temporarily detached modular camera component (a) calibrates the projector unit by mimicking the observers' target perspective so that the projector can display (b) a geometry- and color-corrected image on a curtained window.

angle, aspect ratio, and field of view—as well as their extrinsic position and orientation parameters. Although some systems can project geometrically predistorted images for a known observer position onto scanned or modeled nonplanar surfaces beforehand, these surfaces are still fairly simple, such as adjacent even walls. Surfaces with fine geometric details represent overkill for the real-time predistortion realized by processing a high-resolution 3D model.

Until now, all projection surfaces, both planar and nonplanar, have required a uniform white texture. Per-pixel color correction, however, becomes feasible with the enhanced capabilities of recent graphics chips. A projection onto arbitrarily textured surfaces has been achieved, for example, with the aid of a special transparent film material that reflects a portion of the incident light.⁷

Although this technique allows superimposing flat paintings onto projected multimedia content, it cannot be employed to display images onto everyday surfaces for two reasons. The technique

- still requires an artificial transparent canvas that can be applied to a plain surface; and
- it needs a precise, manual pixel-to-pigment registration.

The too-low resolution of today's cameras prevents the implementation of an automated calibration process for this special case.

Color- and geometry-corrected projection onto arbitrarily shaped and textured surfaces is possible in real time, with fully automatic, fast, and robust calibration. A compact device, such as that depicted in Figure 1, has yet to be built, however. Instead, the first proof-of-concept prototypes are a combination of off-the-shelf components—such as a consumer LCD video beamer, a CCD camcorder, and a personal computer with a TV card and a pixel-shading-capable graphics board.

CREATING VIRTUAL PROJECTION CANVASES

The smart projector concept combines camera feedback with structured light projection to gain information about the screen surface and the environment. Calibrating the system does not require having information about either the surface geometry or the internal or external parameters of the projector and camera. This makes the system extremely robust and easy to use—crucial attributes for home-entertainment and similar applications.

The modular camera component can be detached from the smart projector's projection unit for calibration. It must be temporarily placed approximately at the observers' optimal viewing location or *sweet spot*—pointing at the screen surface as Figure 1a shows. The projection unit can be placed at an arbitrary location. Its light frustum must also cover the screen surface area.

During calibration, the camera mimics the *target perspective*—the optimal viewing position for which the projection unit will be calibrated. The user can either define the display area by sketching the outlines of a virtual projection canvas over a portion of the camera image or derive it automatically from the margins of the camera's field of view.

The calibration process compensates for camera lens distortion at the start to provide video images without radial distortion. The system then determines all parameters required for real-time *geometric predistortion* and *color correction* of video frames delivered by a PAL/NTCS-compliant device such as a DVD player or game console. The fully automated calibration process takes less than 30 seconds with the chosen hardware configuration.

After the system has been calibrated, the camera module can be removed. Henceforth, the projector unit corrects incoming video signals geometrically and photometrically in real time at no less than 100 frames per second. If the system projects the corrected images onto the nontrivial screen surface, the observer will see them as they would appear if

For consumer applications, the calibration of smart projectors must be fully automatic, fast, and robust.

projected onto a plain white canvas. However, this projection canvas is completely virtual and does not exist in material reality, as Figure 1b shows.

VANISHING SHAPES

Smart projector system developers also seek to geometrically predistort the input images so that if they are projected onto a geometrically nontrivial surface and observed from an area close to or at the target perspective, these images appear correct. For consumer applications, the calibration of smart projectors must be fully automatic, fast, and robust.

Projection systems sometimes apply wide-field-of-view cameras in a sweet spot position to calibrate multiple overlapping projectors.⁸ Projecting pixels and capturing them with the camera results in a projector-to-camera pixel mapping. For performance reasons, only a subset of projector pixels are usually displayed and captured, while the mapping for the remaining ones are interpolated linearly. For arbitrarily shaped surfaces with fine geometric details, however, a high-resolution pixel correspondence must be generated in an acceptable time.

To realize this goal, developers can adapt time-multiplexed line-strip scanning techniques from structured 3D range-finder systems. Placing the camera at the observer's sweet spot matches its view to the target perspective. Then the developer can apply a variation of a column-row coded-pattern projection methods,⁹ with phase shifting similar to that proposed by Jens Gühring,¹ to compute a *pixel displacement map*. This creates a lookup table that maps *every* camera pixel to the corresponding projector pixel.

The projector unit displays particular calibration images—for example, the line strips for geometry or a uniform white image for photometry. The camera must capture these images while the projector is displaying them. This means that both must be synchronized and for this we must know the camera's latency in terms of projecting the calibration images long enough. This knowledge helps ensure correct synchronization between projection and capturing during the scanning process.

Hardware image compression and data transfer between the camera and receiving device cause latency. This latency is particularly high for consumer camcorders because they do not target real-time image-processing applications. The smart projector determines the camera's latency automatically at the beginning of the geometric cali-

bration process. The projector does this by sending out sample patterns and measuring the maximum time until it can detect these patterns in the recorded camera images.

For an XGA projector resolution of $1,024 \times 768$, a PAL camera resolution of 720×576 , a maximum camera latency of 80 ms for a consumer camcorder delivering an s-video signal over a TV-in channel, and an average image processing duration of 150 ms, the smart projector's total time to generate the displacement map is approximately 28 seconds. The process requires no user intervention.

The different camera and projector resolutions, and their varying distances and perspectives to the screen surface, prevent the displacement map from representing a one-to-one pixel mapping. The mapping might not even be complete because surface portions can lie in shadow areas. Applying multiple projector units can overcome this problem.

Different projected line strips might project on the same camera pixel. To achieve subpixel precision, the system computes and then stores the average values in the displacement map.

If the camera and projector can be calibrated precisely, the system could use the pixel correspondences in the displacement map and triangulation to recover the screen surface's entire 3D geometry. Some 3D scanners function in exactly this way. Given that we do not expect both devices to be located at known positions, the displacement map allows only the mapping of each camera pixel from the target perspective into the projector's perspective. This results in an undistorted perspective even if the screen surface's 3D shape is unknown.

To benefit from hardware-accelerated computer graphics, the displacement map is converted into a texture map—realized with a 32-bit/16-bit P-buffer—that stores a reference for *every* projector pixel and its corresponding video and camera pixels. The system then passes this texture as a parameter to a modern *pixel shader*, which implements real-time image warping via a *pixel displacement mapping*. Standard components today in many consumer graphics cards, pixel shaders enable per-pixel operations. Besides the *displacement texture map*, the system passes several other parameter textures to the pixel shader, such as the uncorrected input image itself.

To execute a geometric-image predistortion, the system need only render a single 2D rectangle into the projector's entire frame buffer. This triggers the rasterization of every projector pixel through the pixel shader before display. The colors of incom-



Figure 2. Projecting images onto environmental surfaces. (a) A scruffy corner serves as the projection surface. (b) The uncorrected image. (c) The projector system corrects the image's geometry and, finally, (d) its color. Displayed content: The Jackal, Universal Pictures.

ing pixels are simply overwritten by new colors that result from the corresponding input image pixels. These input image pixels can be found with the aid of the displacement texture map, which has the same effect as actually moving the input image pixels to new positions within the projector frame buffer. The colors are not just copied from input image pixels to projector pixels—they are also modified to enable color correction. This allows warping every pixel of the input image in real time without first acquiring geometric information of the screen's surface.

Viewers can perceive the image as geometrically correct at or near the target perspective. Depending on the screen surface's shape, the observer will detect a more or less extreme distortion. For the walls in Figure 2, a horizontal deviation from the target perspective leads to a larger distortion than a vertical deviation, yet the pitched roof surface in Figure 3 would cause the opposite effect. The latter surface is thus better suited to a group of users sitting or standing next to each other.

These are two extreme cases, however. The geometric distortion of, for example, a window curtain and a natural stone wall are relatively low in any direction if viewers are observing the projected image from an adequate distance.

Even though viewers observe the predistorted projected image as geometrically correct on a non-trivially shaped surface, its uncorrected colors will still be blended with the screen surface's texture.

NEUTRALIZED TEXTURES

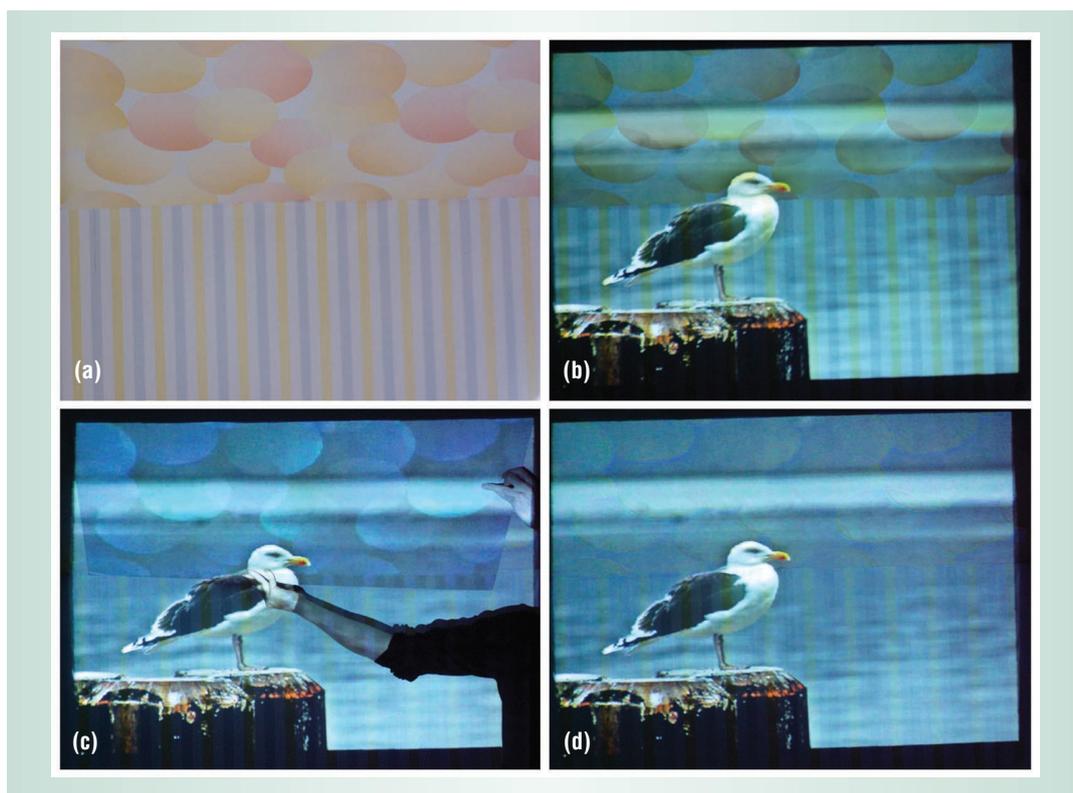
When light strikes a surface, only a fraction of its original intensity and color reflects back; the surface absorbs the rest. For Lambertian (completely diffuse) surfaces, the amount and color of reflected light depends on several parameters, such as the surface's material color (M), the light color and intensity that leaves the source (I), as well as the distance (r) and the incidence angle (α) of light rays with respect to the surface—together called the *form factor* (F). For perfectly diffuse surfaces, Lambert's law approximates the diffuse reflection of light for each spectral component with $R = IFM$, where $F = \cos(\alpha)/r^2$.

In addition to the light that a video projector projects, the environment light is color blended with the surface texture in the same way. Assuming additive color mixing, we can extend Lambert's law to take this into account: $R = EM + IFM$, where E is the environmental light's intensity and color. Environmental light differs from projected light in that the latter can be controlled.

Smart projectors seek to neutralize this natural blending effect by projecting an image (I) in such a way that its blended version on the screen surface appears to observers in its known original colors (R). Given that we consider only diffuse Lambertian screen surfaces—most other surface types are improper for a video projection—we must simply solve the equation for $I:I = (R - EM)/FM$.

Because we do not require information about the

Figure 3. Projection onto a pitched roof area. The image sequence shows (a) the wallpapered surface, (b) the projection with uncorrected colors, (c) color correction projected onto a white piece of paper, and (d) the color-corrected image on wallpaper. All projections are geometry corrected.



projector's or camera's internal and external parameters, we cannot determine each E , M , and F component individually. Rather, we can measure the products EM and FM while R is the given input image—a video frame, for example.

If the video projector displays a bright white image ($I = 1$) onto the screen surface within a dark environment ($E = 0$), the camera captures an image proportional to FM . Further, if we turn off the projector ($I = 0$), the screen surface image captured under environmental light is proportional to EM . These assumptions imply a color- and intensity-adjusted projector and camera, with automatic brightness control, focus, and white-balancing turned off. These simple approximations let us determine the required parameters robustly, without performing complicated measurements or using additional special-purpose devices.

We can use modern *pixel shader* hardware to perform the final correction computations in real time—no less than 100 fps on an Nvidia GeForce FX6800GT—and represent all these parameters as textures. Today, many consumer graphics cards use pixel shaders as standard components and allow per-pixel operations.

The system warps pixels of images taken from the camera view (EM and FM), as well as the input image R , to the projector view via pixel-displacement mapping. This ensures a correct concatenation of corresponding pixels. The resulting image I is finally displayed from the perspective of the projector. These computations are performed on all three

RGB color channels separately. In addition, the pixel shader allows fine-tuning of the output images by considering manually set color, brightness, and gamma correction parameters. It also clips out extreme-intensity situations to avoid visible artifacts.

Figure 3 shows an example of a geometry-corrected projection onto a wallpapered, pitched-roof area. If the input image is not color corrected, the projected colors blend with the colors of the screen surface, as Figure 3b shows.

The wallpaper texture interferes with the video image, which results in a disturbing effect. The color-corrected image (I) is partially shown in Figure 3c by projecting it onto white cardboard. Blending I with the screen surface results in the image shown in Figure 3d, which closely approximates the original input image R . In this case, the screen surface becomes almost invisible. All figures show freeze images of movie frames that the smart projector normally corrects continuously, on the fly, during playback.

BEYOND THE MEANS

Obviously, both geometry correction and color correction will fail if the screen surface's material absorbs the light entirely. Failure will also occur if the surface completely absorbs even part of the spectrum that is visible to the camera. Fortunately, absorbent materials such as velvet are comparatively uncommon in everyday environments, and most diffuse materials produce fairly acceptable results. If the surface can reflect a certain fraction



Figure 4. Image projection onto a curtain. (a) The display surface, a checkered curtain, shows the differences between a projected image with (b) uncorrected colors and (c) one with corrected colors. Both projections are geometry corrected. Displayed content: Finding Nemo, Disney/Pixar Animation Studios.

of the desired light color, the projector only needs to determine how much incident light it must project to produce the desired output. If one projector cannot generate the necessary light, multiple projectors can do so by complementing each other.

Additionally, several technical limitations lower the quality of a current smart projector, such as the limited resolution of consumer camcorders that use PAL or NTSC. If it is necessary to place the camera far away from a large screen surface to capture the entire display area, the projector cannot detect and correct fine surface details. Higher-resolution cameras, such as megapixel digital cameras, can provide better-quality images. In fact, most camcorders already combine two devices in one: a high-resolution digital camera and a video camera that delivers a live video stream. This combination facilitates both fast geometry correction and high-quality color correction.

On the projector side, the limited resolution, low dynamic range, small color space, and high black level of consumer devices represent the main restrictions. A too-low projector resolution causes overly large pixel projections that cannot cover smaller pigments on the screen surface precisely. In particular, the inability to control the black level contributes to the environmental light. Even in a completely dark room, a projector's black level causes the screen surface to be visible. As occurs with a normal projection onto a regular canvas, the black level and the environmental light make displaying dark colors difficult.

However, the human visual system adapts well to local contrast effects. Dark areas surrounded by brighter ones appear much darker than they actually are. Even though researchers will solve these problems in future projector generations, one general problem will remain: Their limited depth focus prevents conventional projectors from displaying

images on extremely curved screen surfaces. Because laser projectors, which can focus on non-planar surfaces, remain far too expensive for the consumer market, using multiple projectors offers a promising solution. The realization of a multifocal projection extension is part of our current research efforts.

One issue remains when projecting onto non-planar surfaces: A single projector can cast shadows on the screen surface that, from the target perspective, appear as cuttings in the presented output image. However, other projectors that contribute from different directions can cover these shadow areas, as Figure 4 shows.

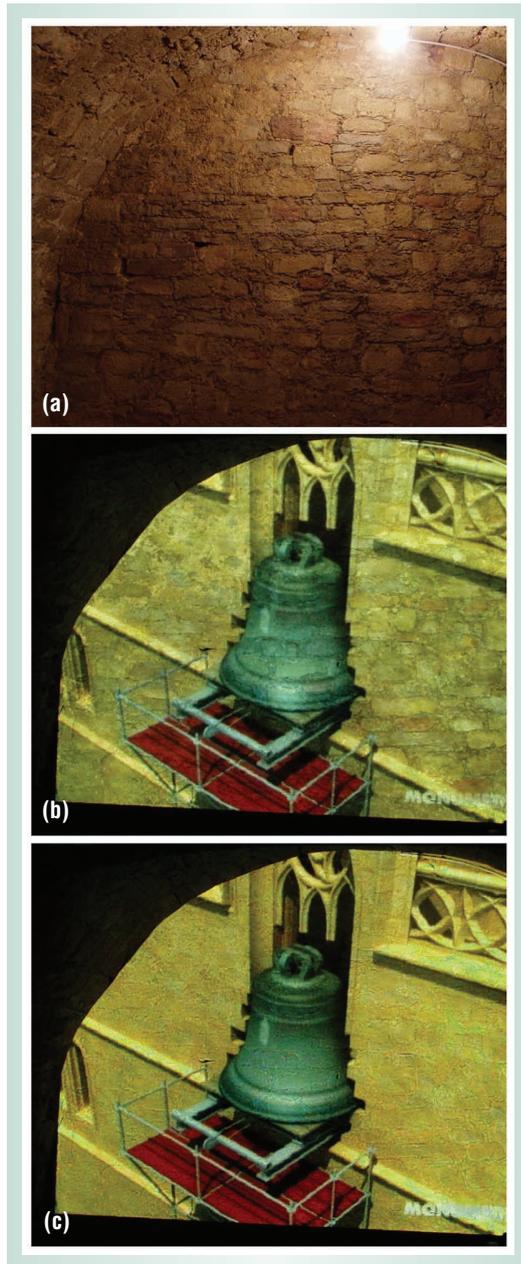
JOINT FORCES

Multiple projectors can enhance the final quality of the output image by complementing each other to achieve increased light intensities and canceling out individual shadow regions. Their output images can fully or partially overlap or can be completely independent. In addition, using multiple projectors allows covering large screen surfaces with high-resolution image tiles. These configurations, known as *tiled screen displays*, provide an overall resolution that a single projector cannot achieve.

In the example in Figure 2, two partially overlapping projectors generate a high-resolution 16:9 format. A smart projector must be scalable so that the end user's calibration effort does not increase with the number of applied projector units. As with the single projector configuration, multiple projectors can be aligned arbitrarily—the calibration process remains fully automatic.

Realizing a second proof-of-concept prototype with two projector units required using a dual-output graphics card to synchronize them. During the geometry calibration, the system generates two dis-

Figure 5. (a) Stenciled projection onto a natural stone wall inside a castle vault. Both projections have undergone geometry correction, while (b) is color uncorrected and (c) is color corrected. Displayed content: The Recovery of Gloriosa, Bennert-Monumedia GmbH.



placement maps sequentially, one for each projector. Consequently, the graphics card can map pixels from the camera view into each projector's perspective so that all pixels display at exactly the same spot on the screen surface. Thus, for N projectors the individual light intensities add up to $R = EM + I_1F_1M + I_2F_2M + \dots + I_NF_NM$.

We can achieve a balanced load among all projectors by assuming that $I_i = I_1 = I_2 = \dots = I_N$. This implies that $R = EM + I_i(F_1M + F_2M + \dots + F_NM)$, and we can solve for $I_i = (R - EM)/(F_1M + F_2M + \dots + F_NM)$. This is equivalent to the assumption that a single high-capacity projector produces the total intensity arriving on the screen surface virtually. Physically, however, the intensity is evenly distributed among multiple low-capacity units.

Although each projector sends the same output intensity, the potentially varying form factors cause different fractions to arrive at the screen surface. The smart projector mixes these fractions on the surface, leading to the final result of $R = EM + I_1F_1M + \dots + I_iF_iM + \dots + I_NF_NM = EM + (R - EM)(F_1M + \dots + F_NM)/(F_1M + \dots + F_NM) = R$.

As for a single projector, a pixel shader that receives the parameter textures EM , and $F_1M \dots F_NM$ computes I_i in real-time. The form factor components F_iM can be determined in two ways:

- by sequentially sending out a white image ($I = 1$) from each projector and capturing each component one by one, or
- by capturing a single image proportional to $F_1M + \dots + F_NM$ by sending each projector's maximum contribution simultaneously.

Although the second method is conceptually more compact, the first method prevents the system from overmodulating the camera's CCD/CMOS sensor. The system captures shadow regions that individual projectors cause in the form factor components. Consequently, the system cancels out the shadows that individual projector units automatically create as a side effect. This, however, implies that the projectors are placed so that at least one projector can reach each surface portion. The cross-fading techniques common to multiprojector setups can achieve smooth transitions among different contributions.

The total duration for calibration increases linearly with the number of projectors. Thus, the two-projector setup can be geometry- and color-calibrated in less than one minute.

Figure 5 shows that, by using multiple smart projectors, we can project images onto surfaces that are neither plain nor white and need not have a rectangular shape. Instead, we can convert many existing surfaces to a display screen by projecting color- and geometry-corrected images onto them. In the consumer context, this capability offers the advantage of fast, fully automatic, and robust calibration, and it allows the correction of video signals in real time. It isn't necessary to know either geometry information or projector and camera parameters. Instead, the projector system performs the entire calibration and correction on a per-pixel level.

Video projectors will play a major role in future home entertainment and edutainment applications—ranging from movies and television

to computer games and multimedia presentations. Smart video projectors have the potential of sensing the environment and adapting to it. This promotes a seamless embedding of display technology into our everyday life.

Future hardware improvements will pave the way for further smart-projector advancements. Upcoming graphics chips, for instance, will be more powerful than ever before. Successive projector generations will continue to feature enhanced quality factors such as brightness, resolution, dynamic range, and black-level. Simultaneously, prices will continue to drop as the smart-projector market share increases. The development of digital cameras and camcorders will follow a similar pattern. Thanks to ongoing miniaturization, all these components could soon be integrated into compact and mobile devices as inconspicuous as light bulbs.

To make immersive 3D visualizations possible within arbitrary environments, we are currently extending the smart-projector concept toward large-scale stereoscopic and multifocal projection. When hardware and software developments have improved, converting a bookshelf into a TV screen or turning a child's entire room into an interactive virtual playground could become possible. ■

Acknowledgments

We thank Gordon Wetzstein and Anselm Grundhöfer for their help on the pixel shader implementation.

References

1. J. Gühring, "Dense 3D Surface Acquisition by Structured Light Using Off-the-Shelf Components," *Proc. Videometrics and Optical Methods for 3D Shape Measuring*, SPIE, 2001, vol. 4309, pp. 220-231.
2. R. Raskar, G. Welch, and H. Fuchs, "Seamless Projection Overlaps Using Image Warping and Intensity Blending," *Proc. 4th Int'l Conf. Virtual Systems and Multimedia*, IEEE Press, 1998, www.cs.unc.edu/~fuchs/publications/VSMm_seamless.pdf.
3. R. Yang et al., "PixelFlex: A Reconfigurable Multi-Projector Display System," *Proc. IEEE Conf. Visualization*, IEEE Press, 2001, pp. 68-75.
4. A. Majumder et al., "Color Calibration of Projectors for Large Tiled Displays," *Proc. IEEE Conf. Visualization*, IEEE Press, 2000, pp. 102-108.
5. R. Raskar et al., "iLamps: Geometrically Aware and Self-Configuring Projectors," *Proc. ACM Siggraph*, ACM Press, 2003, pp. 809-818.
6. R. Raskar et al., "Multiprojector Displays Using Camera-Based Registration," *Proc. IEEE Conf. Visualization*, IEEE Press, 1999, pp. 161-168.
7. O. Bimber et al., "Superimposing Pictorial Artwork with Projected Imagery," to appear in *IEEE Multi-Media*, Jan.-Mar. 2005.
8. R. Sukthankar, T.J. Cham, and G. Sukthankar, "Dynamic Shadow Elimination for Multi-Projector Displays," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, IEEE CS Press, 2001, pp. 151-157.
9. R. Sukthankar, R. Stockton, and M. Mullin, "Automatic Keystone Correction for Camera-Assisted Presentation Interfaces," *Proc. Int'l Conf. Multimodal Interfaces*; http://www.ri.cmu.edu/pubs/pub_3396.html.

Oliver Bimber is a junior professor of Augmented Reality at Bauhaus University, Weimar, Germany. His research interests include display technologies, computer graphics, and human-computer interaction. Bimber received a PhD in engineering from Darmstadt University of Technology, Germany. Contact him at obimber@computer.org.

Andreas Emmerling is a graduate student at Bauhaus University. His research interests include projection technology, computer graphics, and computer vision. Contact him at andreas.emmerling@medien.uni-weimar.de.

Thomas Klemmer is a graduate student at Bauhaus University. His research interests include augmented reality, robotics, and real-time rendering. Contact him at thomas.klemmer@medien.uni-weimar.de.

Get access

**to individual
IEEE Computer Society
documents online.**

More than 100,000
articles and conference
papers available!

\$9US per article for members

\$19US for nonmembers

[www.computer.org/
publications/dlib](http://www.computer.org/publications/dlib)

Interacting with augmented holograms

Oliver Bimber*, Thomas Zeidler, Anselm Grundhoefer, Gordon Wetzstein,
Mathias Moehring, Sebastian Knoedel, and Uwe Hahne
Bauhaus-University Weimar, Media Faculty, Bauhausstr. 11, 99423 Weimar, Germany

ABSTRACT

Holography and computer graphics are being used as tools to solve individual research, engineering, and presentation problems within several domains. Up until today, however, these tools have been applied separately. Our intention is to combine both technologies to create a powerful tool for science, industry and education. We are currently investigating the possibility of integrating computer generated graphics and holograms.

This paper gives an overview over our latest results. It presents several applications of interaction techniques to graphically enhanced holograms and gives a first glance on a novel method that reconstructs depth from optical holograms.

Keywords: Computer Graphics, Human-Computer Interaction, Projection Technology, Real-Time Rendering, Augmented Holograms, Depth Reconstruction.

1. INTRODUCTION

Among all imaging techniques that have been invented throughout the last decades, computer graphics is one of the most successful tools today. Many areas in science, entertainment, education, and engineering would be unimaginable without the aid of 2D or 3D computer graphics. The reason for this success story might be its interactivity, which is an important property that is still not provided efficiently by competing technologies – such as holography.

While optical holography and digital holography are limited to presenting a non-interactive content, electroholography facilitates the computer-based generation and display of holograms at interactive rates^{2,3,29}. Holographic fringes can be computed by either rendering multiple perspective images, then combining them into a stereogram⁴, or simulating the optical interference and calculating the interference pattern⁵. Once computed, such a system dynamically visualizes the fringes with a holographic display. Since creating an electrohologram requires processing, transmitting, and storing a massive amount of data, today's computer technology still sets the limits for electroholography. To overcome some of these performance issues, advanced reduction and compression methods have been developed that create truly interactive electroholograms. Unfortunately, most of these holograms are relatively small, low resolution, and cover only a small color spectrum. However, recent advances in consumer graphics hardware may reveal potential acceleration possibilities that can overcome these limitations⁶.

In parallel to the development of computer graphics and despite their non-interactivity, optical and digital holography have created new fields, including interferometry, copy protection, data storage, holographic optical elements, and display holograms. Especially display holography has conquered several application domains. Museum exhibits often use optical holograms because they can present 3D objects with almost no loss in visual quality. In contrast to stereoscopic or autostereoscopic graphics displays, holographic images can provide all depth cues—perspective, binocular disparity, motion parallax, convergence, and accommodation—and theoretically can be viewed simultaneously from an unlimited number of positions. Displaying artifacts virtually removes the need to build physical replicas of the original objects. In addition, optical holograms can be used to make engineering, medical, dental, archaeological, and other recordings—for teaching, training, experimentation and documentation. Archaeologists, for example, use optical holograms to archive and investigate ancient artifacts^{7,8}. Scientists can use hologram copies to perform their research without having access to the original artifacts or settling for inaccurate replicas.

* bimber@uni-weimar.de; phone +49-(0)3643-583724; fax +49-(0)3643-583724; www.uni-weimar.de/medien/AR;
www.holographics.de (**high-resolution color versions of images presented in this paper are available online**)

Optical holograms can store a massive amount of information on a thin holographic emulsion. This technology can record and reconstruct a 3D scene with almost no loss in quality. Natural color holographic silver halide emulsion with grain sizes of 8 nm is today's state-of-the-art¹⁴.

Today, computer graphics and raster displays offer a megapixel resolution and the interactive rendering of megabytes of data. Optical holograms, however, provide a terapixel resolution and are able to present an information content in the range of terabytes in real-time. Both are dimensions that will not be reached by computer graphics and conventional displays within the next years – even if Moore's law proves to hold in future.

Obviously, one has to make a decision between interactivity and quality when choosing a display technology for a particular application. While some applications require high visual realism and real-time presentation (that cannot be provided by computer graphics), others depend on user interaction (which is not possible with optical and digital holograms). Consequently, holography and computer graphics are being used as tools to solve individual research, engineering, and presentation problems within several domains. Up until today, however, these tools have been applied separately. Our intention is to combine both technologies to create a powerful tool for science, industry and education. We are currently investigating the possibility of integrating computer generated graphics and holograms¹. Our goal is to combine the advantages of conventional holograms (i.e. extremely high visual quality and realism, support for all depth queues and for multiple observers at no computational cost, space efficiency, etc.) with the advantages of today's computer graphics capabilities (i.e. interactivity, real-time rendering, simulation and animation, stereoscopic and autostereoscopic presentation, etc.).

The remainder of this paper is organized as follows: Chapter 2 briefly summarizes the general concept of digitizing the reconstruction wave to reply a hologram. Spatial distribution, amplitude and wavelength can be controlled to augment holograms consistently with graphical elements. Since depth information of both –holographic and graphical content– is essential for our approach, chapter 3 presents our initial results of using a flat bed scanner for estimating holographic depth information. Chapter 4 discusses several interaction forms in combination with different hologram types and describes current experimental display configurations. It is shown how the viewing range of augmented holograms can be increased with the aid of multiplex stereograms and how augmented holograms can be embedded into window-oriented desktop workplaces. Furthermore, we demonstrated how force-feedback and touch-interaction can be applied to augmented holograms, and how to combine volumetric multiplexed holograms with stereoscopic graphics. Finally, chapter 5 concludes our contribution and identifies some limitations and future directions.

2. DIGITAL LIGHT FOR RECONSTRUCTING HOLOGRAMS

The two basic hologram types—transmission and reflection—are both reconstructed by illuminating them with spatially coherent light (i.e. using a point-source of light). These two types have generated a number of variations. Although some holograms can be reconstructed only with laser light, others can be viewed under white light.

Conventional video projectors represent point sources that are well suited for viewing white-light reflection or transmission holograms. Early experiments with video projectors for reconstructing optical holograms have been made in the art and engineering domains. In some art installations, optical holograms have been linked with time-based media, such as slides, film-loops or color effects that are projected onto them to achieve artistic effects^{15, 16}. Others have redirected projected light with multiple mirrors to simulate multiple different light sources. The goal was to achieve dynamic fluctuation effects with optical holograms^{17, 18}.

Today's high-intensity discharge lamps of projectors can produce a very bright light. The main advantage for using video projectors instead of analog light bulbs is that the reference wave used to reconstruct the hologram can be digitized. Thus it is possible to control the amplitude and wavelength of each discrete portion of the wavefront over time¹. Today's computer graphics capabilities allow merging any kind of two-dimensional or three-dimensional graphical elements seamlessly with the recorded holographic content – potentially leading to efficient visualization tools that combine the advantages of holography and computer graphics.

A hybrid display approach has already been described earlier that combined a transmission hologram with a liquid crystal display to realize a new user interface for business machines, such as photocopiers¹⁹. In this case, a normal light bulb was used to illuminate a transmission hologram which was mounted behind an LCD panel. Since the light source was analog, it was not possible to control the reconstruction of the holographic content. Electronically activated two-dimensional icons that were displayed on the LCD appeared simultaneously and unregistered with the three-dimensional hologram. As mentioned above, video projectors allow digitizing the reference wave, which leads to a seamless integration of graphical elements into an optical hologram.

Figure 1 shows the projected reference wave (top row) in different states, and the resulting holographic image (bottom row) of a monochrome white-light reflection hologram. A uniform reference wave reconstructs the entire hologram uniformly (first column of fig. 1). Selectively emitting light in different directions allows us to create an incomplete reference wave that reconstructs the hologram only partially (second column of fig. 1). Local amplitude variations in the reference wave result in proportional amplitude variations in the reconstructed object wave (third column of fig. 1). Variations in wavelength do not lead to useful effects in most cases due to the wavelength dependency of holograms (fourth column of fig. 1). But this is still a matter for further investigations. One example for encoding color information into the reference wave is described in section 4.6.

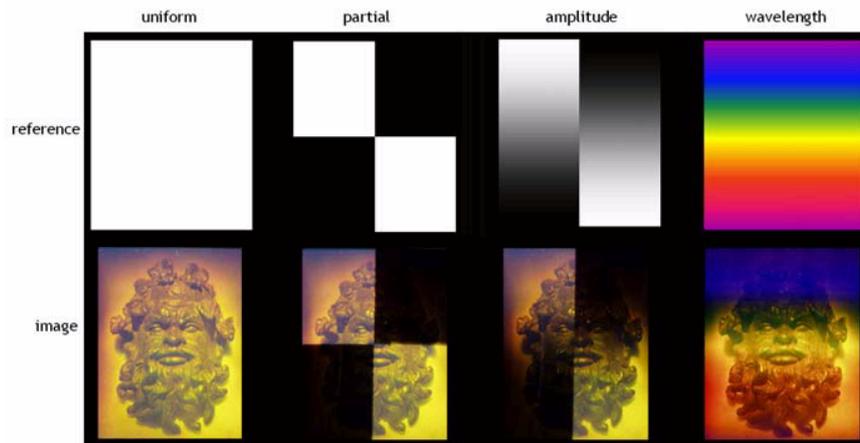


Figure 1: The projected reference waves (top row) and the resulting holographic images (bottom row).

The following sub-sections will give only a brief overview of the possibilities of using a digitized reference wave for reconstructing optical holograms. The interested reader is referred to the primary publication¹ for details on rendering and illumination techniques.

2.1. Partially reconstructing object waves

It is possible to reconstruct the object wave of a hologram only partially, leaving gaps where graphical elements can be inserted. Both reflection holograms (without an opaque backing layer) and transmission holograms remain transparent if not illuminated. Thus, they can serve as optical combiners—leading to very compact displays. Real-time computer graphics can be integrated into the hologram from one side, while illuminating it partially from the other side¹.

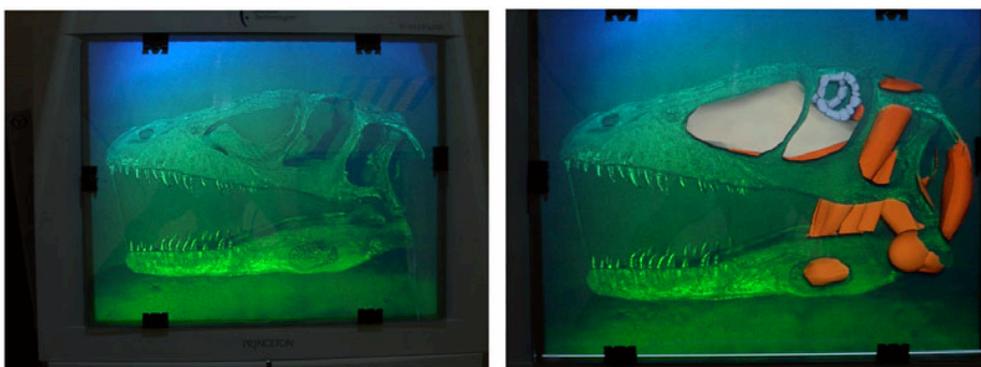


Figure 2: Rainbow hologram of a dinosaur skull combined with graphical representations of soft tissue and bones.

Thereby, rendering and illumination are view-dependent and have to be synchronized. If autostereoscopic displays are used to render 3D graphics registered to the hologram, both holographic and graphical content appear three-dimensional

within the same space. If depth information of both is known, correct occlusion effects between hologram and graphics can be generated. One possible approach of acquiring depth information directly from a white-light hologram is explained in chapter 3.

Figure 2 shows a rainbow hologram of a dinosaur skull combined with graphical representations of soft tissue and bones. If the holographic plate is illuminated with a uniform light, the entire hologram is reconstructed (fig. 2-left). If the plate is illuminated only at the portions not occluded by graphical elements, the synthetic objects can be integrated by displaying them on the screen behind the plate (fig. 2-right). Technical details on how this is achieved will be presented in section 4.1.

2.2. Light interaction

The reconstructed object wave's amplitude is proportional to the reference wave's intensity. In addition to using an incomplete reference wave for reconstructing a fraction of the hologram, intensity variations of the projected light permit local modification of the recorded object wave's amplitude.

Practically, this means that to create the illumination image which is sent out by the projector, graphical shading and shadowing techniques are used to reconstruct the hologram instead of illuminating it with a uniform intensity. To do this, the real shading effects on the captured scenery caused by the real light sources used for illumination during hologram recording, as well as the physical lighting effects caused by the video projector on the holographic plate, must both be neutralized. Next, the influence of a synthetic illumination must be simulated¹.

Using conventional graphics hardware, it becomes possible not only to create consistent shading effects, but also to cast synthetic shadows correctly from all holographic and graphical elements onto all other elements.



Figure 3: A rainbow hologram with 3D graphical elements and synthetic shading and shadow effects.

Figure 3 shows the same rainbow hologram as above with 3D graphical elements and synthetic shading effects. Shadows are cast correctly from the hologram onto the graphics and vice versa. A virtual point-source of light was first located at the top-left corner (L_1 in fig. 3-left), and then moved to the top-right corner (L_2 in fig. 3-right), in front of the display. Moving the virtual light source and computing new shading effects can be done in real-time. Note, that only intensity/shading variations are simulated in figure 3. A vertical variation of the object wave's wavelength that is due to diffraction effects of the rainbow hologram is still visible and cannot be corrected. This is not the case for white-light reflection holograms.

3. RECONSTRUCTING HOLOGRAPHIC DEPTH WITH FLATBED SCANNERS

Depth information of the recorded holographic content is essential for correct rendering and interaction in combination with augmented graphical elements. The shading and occlusion effects that are presented in sections 2.1 and 2.2, as well as the force-feedback simulation that is discussed in section 4.4 would not be possible without knowing the surface geometry of the holographic content. For demonstrating the capabilities of the various techniques, the object's surface geometry has been laser-scanned before recording it as a hologram. The scanned geometric model is then registered to its holographic counterpart during a pre-process. This allows computing estimated depth values of the surfaces recorded

in the hologram. However, since the original objects are usually not available for scanning after the recording process, the depth information has to be extracted directly out of the hologram.

Ultra-fast holographic cameras, for instance, have been modified to allow capturing 3D objects, such as faces²⁰ or bodies²¹. A fast pulsed laser with short exposure time ($25ns$) is used in these cases for holographic recording that is free of motion artifacts. The depth information is then reconstructed by illuminating the hologram with a laser. Topometric information is retrieved by digitizing the real holographic image that is projected onto a diffuse plate. Moving the plate in the depth direction (away from the holographic plate) results in several 2D slices through the holographic image. These slices are finally combined to form the corresponding 3D surface.

Another approach measures the shape of a recorded surface by determining the time for light to travel from different points of the object²². They are based on the holographic light-in-flight technique²³.

The sections below describe our initial results of a new approach that uses a flat bed scanner for estimating holographic depth information.

3.1 Scanning Multiple Views

Two commercial types of flatbed scanners exist today: Most scanners apply a CCD (charge-coupled device) array that captures a parallel projected image in the moving direction of the scanning slit and a perspective projected image in the other direction. A new optical technology has been introduced by Canon –called LIDE (LED InDirect Exposure)– to reduce the size of scanners. An array of parallel rod lenses over the entire scanner width creates a parallel projection in both directions. Thus, a LIDE scanner represents a parallel perspective camera with a low focal depth.

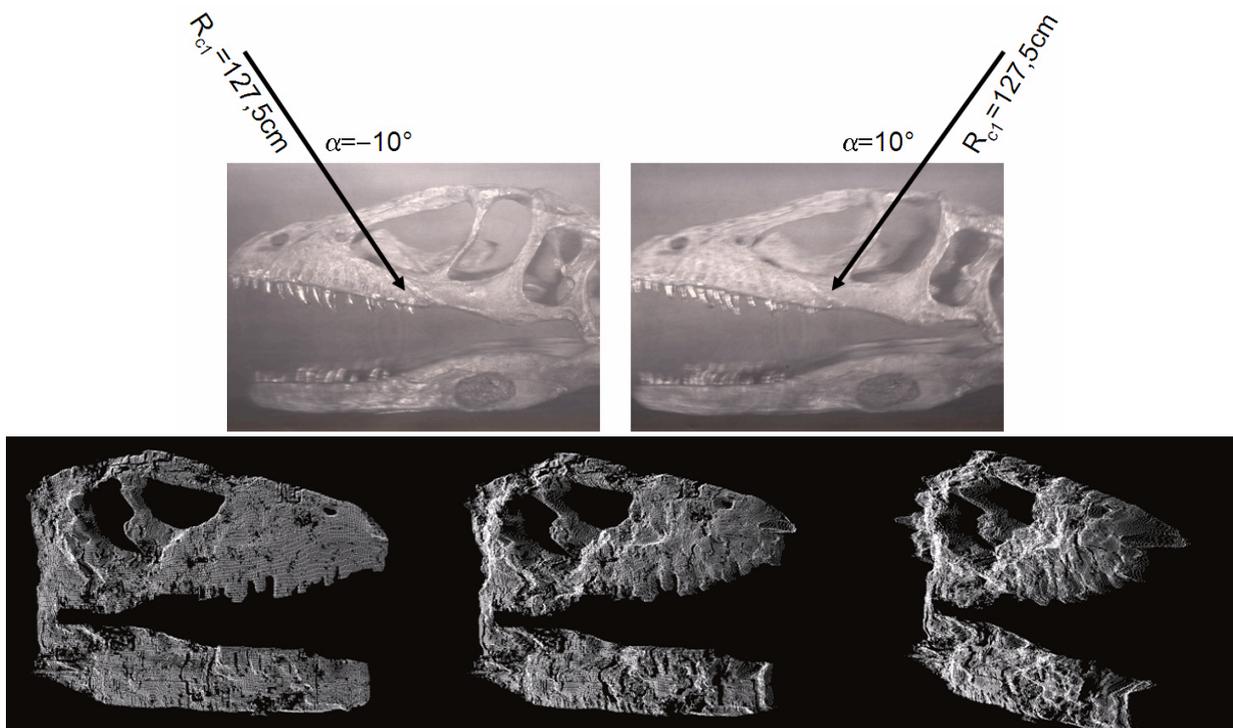


Figure 4: Two LIDE scans of a rainbow hologram with different reference wave angles (top), and point-cloud of reconstructed depth map (bottom).

We use the LIDE technology to scan multiple images of a hologram by placing the holographic film on top of the scanner window, leaving the lit open and illuminating it under different illumination angles for each scan (cf. figure 4). The geometric image distortion that is caused by the different illuminations²⁶ is captured with a parallel camera model defined by the scanner. In this case, an analytical solution exists for computing depth information if the correspondences between 2D image projections (disparities) are known.

3.2 Computing 2D Disparities

Common 3D scanning techniques analyze images that are taken from cameras located on the same base-line by searching for pixel correspondences along a single direction. Due to the complex warping behavior of holograms when they are illuminated under different situations²⁶ our method requires to compute 2D disparities. We apply an algorithm that uses a hierarchical stereo matching strategy using the discrete wavelet transform (DWT)²⁸. As mentioned above, the hologram is reconstructed from at least two different (but known) light positions. Images of the hologram are scanned with a LIDE flatbed scanner while being illuminated from these positions. If the correspondences between a minimum of two images are known it is possible to estimate depth values. More images that are recorded under additional light positions can add redundancy and consequently enhance the outcome. However, the cross-correspondence between single image pairs has still to be computed. Thus we only describe the analytical solution for the basic case of computing depth from two parallel perspective images.

3.3 Estimating Depth

With the computed disparities we aim at estimating the depth z_o of all visible pixels. The geometric imaging behavior of holograms under different lighting conditions has been well understood²⁶. Based on this model, a numerical method has been derived that estimates the image position of recorded objects for cases in which the recording reference beam does not match the reply reference beam, and a known perspective viewing situation is assumed²⁷. This is illustrated in figure 5.

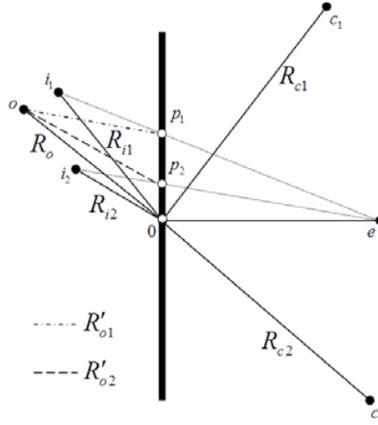


Figure 5: Flatland diagram of reconstruction geometry.

Since the LIDE scanner provides a parallel projection in both directions, an analytical solution can be found. For a parallel projection, the viewpoint e can be assumed to be located at infinity. Because of this the principal ray points p_i are known for both lighting positions c_1 and c_2 : $x_p = x_i$ and $y_p = y_i$. Thus, the reconstructed image points i_i of a recorded object o and the position of the viewer e are collinear. With these two constraints the equation that can be derived for the perspective case²⁷ reduces to:

$$0 = \frac{x_c - x_i}{R'_c} + \mu \left(\frac{x_o - x_i}{R'_o} - \frac{x_r - x_i}{R'_r} \right) \quad (1)$$

This also holds for the other dimension:

$$0 = \frac{y_c - y_i}{R'_c} + \mu \left(\frac{y_o - y_i}{R'_o} - \frac{y_r - y_i}{R'_r} \right) \quad (2)$$

Note, that $\mu = \lambda_c / \lambda_r$ where λ_c is the reconstruction wavelength and λ_r is the recording wavelength. This ratio can vary slightly from the theoretical value due to changes in humidity during reconstruction. Note also, that the radii R

are measured from the origin 0 while the radii R' are measured from the corresponding principle ray points p_i . The origin and the principle ray points lie on the holographic plane. Equations (1) and (2) are valid for all light positions. Solving these equations for x_0 and y_0 we receive:

$$x_0 = R'_{o1} \underbrace{\left(\frac{x_r - x_{i1}}{R'_{r1}} - \frac{x_{c1} - x_{i1}}{\mu R'_{c1}} \right)}_a + x_{i1} \quad \text{and} \quad x_0 = R'_{o2} \underbrace{\left(\frac{x_r - x_{i2}}{R'_{r2}} - \frac{x_{c2} - x_{i2}}{\mu R'_{c2}} \right)}_b + x_{i2} \quad (3)$$

$$y_0 = R'_{o1} \underbrace{\left(\frac{y_r - y_{i1}}{R'_{r1}} - \frac{y_{c1} - y_{i1}}{\mu R'_{c1}} \right)}_c + y_{i1} \quad \text{and} \quad y_0 = R'_{o2} \underbrace{\left(\frac{y_r - y_{i2}}{R'_{r2}} - \frac{y_{c2} - y_{i2}}{\mu R'_{c2}} \right)}_d + y_{i2} \quad (4)$$

From equation (3) follows:

$$R'_{o1} = \frac{bR'_{o2} + x_{i2} - x_{i1}}{a} \quad (5)$$

Equation (4) yields:

$$R'_{o1} = \frac{dR'_{o2} + y_{i2} - y_{i1}}{c} \quad (6)$$

With equations (5) and (6) R'_{o2} can be calculated, and with known R'_{o2} the depth for every projected object o is:

$$z_o = R'_{o2} \sqrt{1 - \frac{(x_o - x_{i2})^2}{R'_{o2}{}^2} - \frac{(y_o - y_{i2})^2}{R'_{o2}{}^2}} \quad (7)$$

Note, that correct disparities (which define the correspondences between the two images i_1 and i_2) are essential for computing the correct depth values. This still represents our main challenge.

4. ADDING INTERACTIVITY

We believe that interactivity is one of the success factors of modern computer graphics. Electroholography has a great potential to provide truly interactive holograms. However, several technological hurdles have to be taken before this will become a real competitor to computer graphics. Combining holograms and interactive computer graphics represents an intermediate solution that can be achieved today with off-the-shelf equipment. The following sections describe current experimental display prototypes and present several interaction forms in combination with different hologram types.

4.1 Display prototypes

Figure 6 shows two desktop prototypes. They serve as proof-of-concept configurations and as testbeds for experiments. The stereoscopic version (figure 6-left) consists of a conventional CRT screen with active stereo glasses, wireless infrared tracking, and a touch screen in front of the hologram for interaction. The autostereoscopic version uses an autostereoscopic lenticular-lens sheet display with integrated head-finder for wireless user tracking and a force feedback device⁹ for six degrees-of-freedom interaction.

For the autostereoscopic display prototype, a digital light projector (DLP) is applied for illuminating the hologram. Since the DLPs' time-multiplexed generation of light intensities causes synchronization conflicts with the shuttering of the active LC glasses, an LCD projector is used for the stereoscopic version instead.

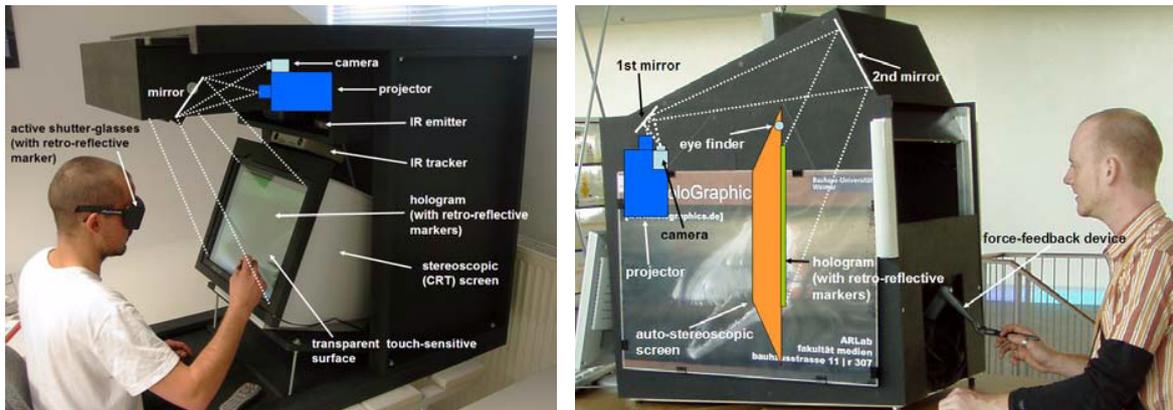


Figure 6: Stereoscopic (left) and autostereoscopic (right) display prototypes.

A single PC with a dual-output graphics card renders the graphical content on the screen and the illumination for the holographic plate on the video projector. In both cases, the screen additionally holds further front layers (cf. figure 7) – glass protection, holographic emulsion, and optional mirror beam combiner (used for transmission holograms only).

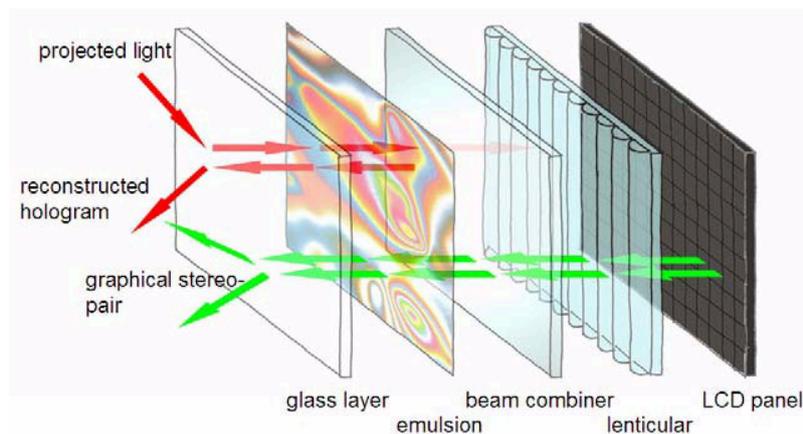


Figure 7: Explosion model of the optical layers' stacked structure. The example shows a transmission hologram in combination with autostereoscopic lenticular screen.

Interaction with the graphical content is supported with a mouse, a touch-sensitive transparent screen mounted in front of the holographic plate, or a 6DOF force feedback device (see section 4.4).

In addition, a camera is mounted close to the projector to detect retro-reflective markers that are attached to the holographic plate. Using structured light probes ensures a fully automatic registration of the holographic plate and calibration of the projector.

4.2 Increasing viewing range with multiplex stereograms

Digital holography uses holographic printers to expose the photometric emulsion with computer-generated or captured images. This results in conventional holograms with digital content rather than real scenery. Pre-processed 2D and 3D graphics or digital photographs and movies can be printed. This allows to holograph, for instance, completely synthetic objects, real outdoor sceneries, and objects in motion - which is difficult and sometimes impossible to achieve with optical holography. Like optical holograms, digital holograms can be multiplexed. This allows to divide the viewing space and to assign individual portions to different contents. The content for digital holograms can easily be created by non-experts, and the printing process is inexpensive. Usually a 3D graphical scene, a series of digital photographs or a

short movie of a real object is sufficient for producing digital holograms. But digital holograms lack in the quality (resolution, color appearance, sharpness, etc.) of optical holograms.

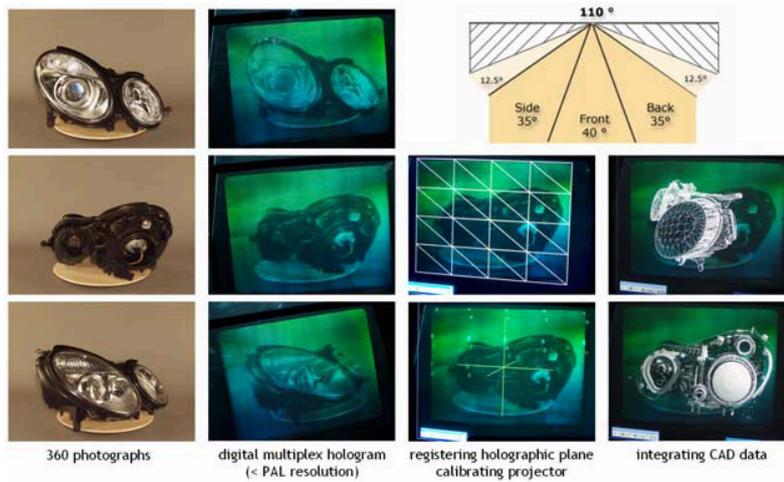


Figure 8: A multiplexed digital reflection stereogram of a car headlight with integrated CAD data.

Figure 8 shows a digital color white-light reflection stereogram of a car's head-light. It was generated by taking 360 perspective photographs from different angles (in 0.5 degree steps to cover a 110° total viewing zone plus two 35° deg clipping areas). The perspective photographs were multiplexed into different sub-zones (40°=80 images for the front view + 2x35°=140 images for the side and rear views + 2x12.5°=50 images to fill the partially visible clipping area outside the 110° total viewing zone + 2x22.5°=90 images to fill the invisible clipping area outside the 110° total viewing zone). Consequently, three different partial views (front, rear, and side) can be observed by moving within the total viewing zone of 110°. After registering the holographic plane and calibrating the projector, interactive graphical elements, such as wire-frame or shaded CAD data can be integrated into the hologram. Since the head motion of the observer is tracked and the recorded viewing angles are known, the perspective of the graphical content can be updated to match the corresponding perspective recorded in the hologram. Thus, graphical and holographic content remain registered – regardless of the observer's viewing direction. This supports a non-continuous surround view of recorded and augmented scenes.

4.3 Holographic windows

The ability to digitally control the reconstruction of a hologram allows integrating them seamlessly into common desktop-window environments. If the holographic emulsion that is mounted in front of a screen is not illuminated, it remains transparent. In this case the entire screen content is visible and an interaction with software applications on the desktop is possible in a familiar way. The holographic content (visible or not) is always located at a fixed spatial position within the screen/desktop reference frame. An application that renders the graphical content does not necessarily need to be displayed in a full screen mode (as in the examples above), but can run in a windows mode - covering an arbitrary area on the desktop behind the emulsion. If position and dimensions of the graphics window are known, the projector-based illumination can be synchronized to bind the projected light to the portion of the emulsion that is located directly on top of the underlying window.

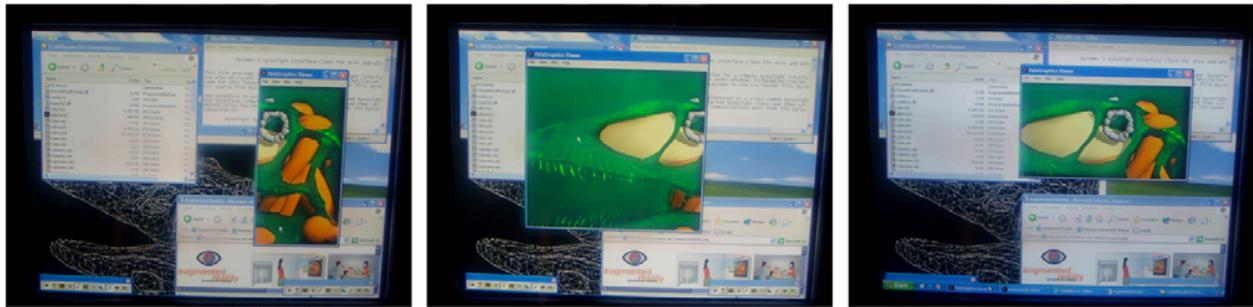


Figure 9: A holographic window in different states on a desktop together with other applications. The hologram is a white-light reflection hologram.

Thereby, all the techniques that are described in section 2 (partial reconstruction and intensity variations) are constrained to the window's boundaries. The remaining portion of the desktop is not influenced by the illumination, the graphical or the holographic content. In addition, the graphical content can be rendered in such a way that it remains registered with the holographic content - even if the graphical window is moved or resized. This simple, but effective technique allows a seamless integration of holograms into common desktop environments. It allows to temporarily minimize the "holographic window" or to align it over the main focus while working with other applications. Figure 9 shows a holographic window in different states on a desktop together with other applications. It displays an optical (monochrome) white-light reflection hologram of a dinosaur skull with integrated graphical 3D soft tissues. A stereoscopic screen was used in this case, because autostereoscopic displays (such as lenticular screens or barrier displays) do not yet allow an undisturbed view on a non-interlaced 2D content (such as text with a small font).

4.4 Force-feedback interaction

Haptic interaction devices, such as Massie's PHANTOM⁹, allow to feel computer generated 3D content by simulating force feedback. Depending on the position of the 6DOF stylus-like device (cf. figures 6-right and 10), force vectors and magnitudes are calculated dynamically based on pre-defined material parameters of the corresponding holographic or graphical content. This enables the user to virtually touch and feel the hologram and the integrated virtual models. For both content types a 3D geometrical model is required to determine surface intersections with the stylus, which lead to the proper force computations. The device is installed outside the user's viewing volume and allows controlling a 3D cursor remotely. This prevents from occluding parts of the screen or casting shadows from the projected light.

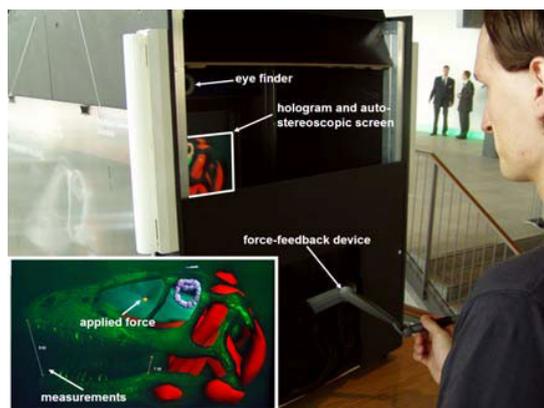


Figure 10: Force-feedback interaction with a reflection hologram of a dinosaur skull and augmented graphical soft-tissue.

While the holographic content is static, the graphical content can be deformed by the device in real time. As in some of the previous examples, a reflection hologram of a dinosaur skull has been augmented with reconstructed soft tissue. Using the force feedback device allows touching and feeling bones and muscles with different material parameters.

While bones feel stiff, muscles and air sinus feel differently soft. In addition, the soft tissue can be pushed in under increasing pressure, and expand back when released. Furthermore, measurements of distances can be taken by touching arbitrary points in space with a virtual measuring tool that is controlled by the stylus.

Other groups have experimented force feedback interaction in combination with static reflection transfer holograms¹⁰, edge-illuminated holograms¹¹, and dynamic electroholograms^{12,13}. Note, that in contrast to electroholograms, optical holograms are static. A modification of the holographic content is not possible in this case. This is also true for our approach. Only the graphical content can be dynamically modified. However, force feedback can be simulated for both - the holographic and the graphical part.

4.5 Touch interaction

A transparent and touch-sensitive surface can be used as front layer to support touch and pointing interactions. The prototype illustrated in figure 11 applies a resistive analog touch screen with an invisible spacer, a touch resolution of 2048x2048 on a surface of 30x40cm, an 80% nominal light transmission, 10ms responds time and a maximum error of 3mm. Touch events are activated through pressure – by finger, fingernail, gloved hand or stylus. The resulting 2D position of the pointer on the registered panel in combination with the head-position of the observer (known from head-tracking) leads to a 3D ray that can be used by ray-casting techniques to select holographic or graphical objects that are intersected by the ray.

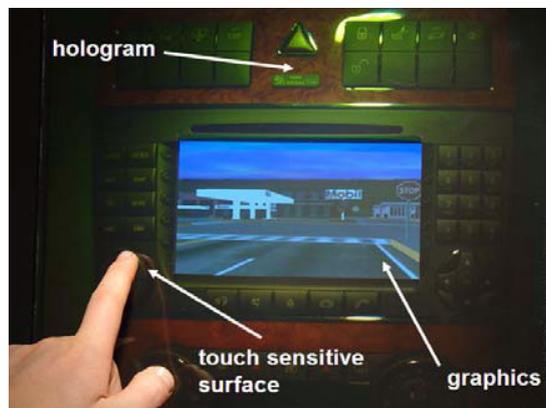


Figure 11: A full color hologram of a car navigation console. Graphics is integrated into the console's display to simulate the design of novel graphics interfaces before building prototypes.

In the example shown in figure 11, touch interaction is used for operating the recorded keypad of a car navigation system. A high-quality full color Denisjuk hologram replays the console while monoscopic or autostereoscopic graphics is integrated into the console's original LCD panel. This allows simulating novel 2D/3D graphics interface designs, as well as new interaction and presentation schemes for navigation systems under realistic evaluation conditions before building physical prototypes. Touching the surface above a recorded button triggers the specific function of the system and the augmented graphics panel is updated with the corresponding content.

4.6 Interacting with volumetric multiplexed holograms

Volumetric multiplexed holograms²⁴ are digital multiple exposure transmission holograms that contain static CT, MRI or other volumetric datasets. They encode slices from a medical scan on a holographic film. The slices can be reconstructed simultaneously and appear as semi-transparent in space – resulting in a three-dimensional volumetric image. In contrast to many other 3D display techniques that are applied for viewing medical data, volumetric multiplexed holograms share the properties of other hologram types and support all depth queues. This is also the case for electroholographic displays. Some experiments have been made recently to use such devices for presenting time-series volumetric data²⁵.

Although the recorded content is static, volumetric multiplexed holograms can be augmented with interactive stereoscopic graphics.

The *Voxbox* display²⁴ is an analog light-box-like display that consists of a small tungsten halogen lamp, a front-surface mirror, and a sandwich of a Fresnel lens, a diffraction grating with dispersion compensation and a light directing film. It

is used to replay volumetric multiplexed holograms. The mirror of the *Voxbox* display virtually places the lamp at the focal point of the lens. The collimated light behind the lens is diffracted at an angle to reconstruct the hologram, and dispersed to cancel its strong dispersion effects. A louver film between the grating and the hologram blocks the zero-order light while letting the grating's dispersed spectrum pass through the hologram.

To integrate stereoscopic computer graphics into the hologram, we replace the halogen lamp and the mirror by two LCD projectors and mirror beam-combiners (cf. figure 12-top). The projected images are used to illuminate the hologram and to display colored stereo pairs at the same time. To ensure that a focused image is projected onto the holographic plane, the original Fresnel lens is replaced by a new lens with a focal distance of $1m$. A 50/50 mirror beam combiner allows locating both projection centers at the lens' focal point. For separating the stereo images, we attach LC shutters in front of each projector's lens. These shutters are triggered in sync with the LC shutters of the observer's glasses by an external pulse generator at $100Hz$ or more. Air cooling prevents the LC shutters from being overheated. A normal (horizontal or vertical) alignment of the projectors would cause diffraction artifacts in form of visible Moiré patterns. This is because the raster of the projected pixels are in line with the orientation of the diffraction grating and the louvers of the light directing film. To minimize these artifacts, both projectors are tilted to a 45° angle around the projection axis. Note that in this case the LC shutters on the observer's glasses have to be rotated by the same angle to prevent light-loss caused by polarization effects. We use a wireless infrared tracking device to support head-tracking of the observer which is required for perspective correct rendering of the graphics for different head positions.

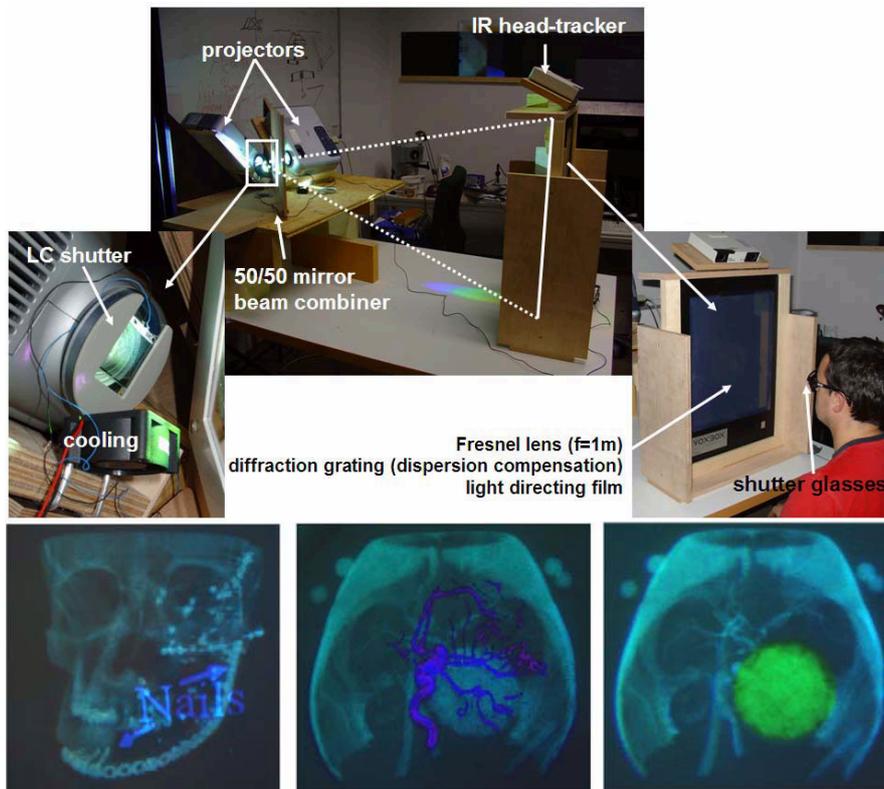


Figure 12: Prototype setup of modified *Voxbox* display (top), and sample results of holographic CT and MRI recordings with integrated stereoscopic graphics (bottom).

One projector displays the left stereo image while the other one displays the right stereo image. If the background of both images is white (or another gray scale), the holographic content is completely reconstructed. At those places where graphical elements are projected onto another color, the holographic content is blended with the graphical content. Thus, the graphical content appears also semi-transparent. A correct occlusion of holographic parts by graphical objects can only be achieved if the graphics is rendered in black (e.g., on a white background). This does not reconstruct the

hologram in these areas. These techniques can be used to integrate interactive graphical augmentations into the hologram, or to color code different parts in a volumetric model, as illustrated in figure 12-bottom. This opens a variety of new interaction possibilities with volumetric multiplexed holograms, which still have to be explored.

5. CONCLUSION

We believe that electroholography will play a major role for future display technology. It has a high potential to provide a platform for a realistic and interactive visualization in many areas. However, several technological problems have yet to be solved before electroholography will be a useful alternative to three-dimensional computer graphics. Holograms support all depth queues. An important property which all other stereoscopic and autostereoscopic displays do not provide. However, non-electronic holograms are static and lack in interactivity.

We have described several rendering, illumination, interaction and reconstruction methods for combining optical or digital holograms with interactive computer graphics. While the holographic content can offer all advantages of holograms (i.e. extremely high visual quality and realism, support for all depth queues at no computational cost, space efficiency, etc.), the integrated graphical part is interactive. As illustrated, a variety of common hologram types, existing displaying methods, and application domains can be addressed with this concept. Computational intensive information, such as volumetric datasets or photorealistic scenes, for instance, can be represented with a hologram while interactive elements can be added with computer graphics. All of this is possible with off-the-shelf technology that is available today. This may lead to new tools for science, industry and education.

Our future work in this area focuses on the exploration of advanced rendering methods, suitable display configurations, useful interaction techniques and improved depth extraction algorithms. Receiving feedback from potential users will allow to refine and validate our concept.

ACKNOWLEDGEMENTS

We would like to acknowledge and thank Yves Gentet of the Yves GENTET - Art and Science holographic studio, Stephen Hart of Voxel Inc., Kaveh Bazargan of Focal Image Ltd, Tim Frieb of Holowood Holographiezentrum Bamberg e.K., Wilhelm Wilke of DaimlerChrysler AG, Holger Regenbrecht of Otago University, Steven Gatesy of Brown University, and Lawrence Witmer of Ohio University for their technical support and professional feedback. This project is supported by the Deutsche Forschungsgemeinschaft (DFG) under contract number BI 835/1-1.

REFERENCES

1. O. Bimber, *Combining Optical Holograms with Interactive Computer Graphics*, IEEE Computer, pp. 85-91, January 2004.
2. J.S. Kollin, S.A. Benton, and M.L. Jepsen, *Real-Time Display of 3-D Computed Holograms by Scanning the Image of an Acousto-Optic Modulator*, Proceedings of SPIE, vol. 1136, Holographic Optics II: Principles and Applications, Springer, pp. 178-185, 1989.
3. M. Lucente, *Interactive Three-Dimensional Holographic Displays: Seeing the Future in Depth*, ACM Computer Graphics, Special Issue on Current, New, and Emerging Display Systems, ACM Press, pp. 63-67, 1997.
4. M. Lucente and A. Tinsley, *Rendering Interactive Images*, Proceedings of ACM Siggraph'95, ACM Press, pp. 387-394, 1995.
5. M. Lucente, *Interactive Computation of Holograms Using a Look-Up Table*, Journal of Electronic Imaging, vol. 2, no. 1, pp. 28-34, 1993.
6. C. Petz and M. Magnor, *Fast Hologram Synthesis for 3D Geometry Models Using Graphics Hardware*, Proceedings of SPIE'03, Practical Holography XVII and Holographic Materials IX, vol. 5005, Springer, pp. 266-275, 2003.
7. F. Dreesen and G. von Bally, *Color Holography in a Single Layer for Documentation and Analysis of Cultural Heritage*, In Proceedings of Optics within Life Sciences (OWLS IV), Springer, pp. 79-82, 1997.
8. F. Dreesen, H. Deléré, and G. von Bally, *High-Resolution Color Holography for Archaeological and Medical Applications*, In Proceedings of Optics within Life Sciences (OWLS V), Springer, pp. 349-352, 2000.
9. T. H. Massie, and J. K. Salisbury, *The PHANTOM Haptic Interface: A Device for Probing Virtual Objects* Proceedings of the ASME Dynamic Systems and Control Division, vol. 55, no. 1, pp. 295-301, 1994.

10. M.R.E. Jones, *The Haptic Hologram*, Proceedings of SPIE, Fifth International Symposium on Display Holography, vol. 2333, pp. 444-447, 1994.
11. W. Plesniak and M. Klug, *Tangible Holography: Adding Synthetic Touch to 3D Display*, Proceedings of the IS&T/SPIE Symposium on Electric Imaging, Practical Holography XI, pp. 53-63, 1999.
12. W. Plesniak and R. Pappu, *Coincident Display Using Haptic and Holographic Video*, Proceedings of ACM SIGCHI conference on Human factors in computing systems, pp. 304-311, 1998.
13. W. Plesniak and R. Pappu, *Spatial Interaction with Haptic Holograms*, Proceedings of the IEEE International Conference on Multimedia Computing and Systems, vol. 1, pp. 413-426, 1999.
14. Y. Gentet and P. Gentet, *Ultimate emulsion and its applications: a laboratory-made silver halide emulsion of optimized quality for monochromatic pulsed and full-color holography*, Proceedings of SPIE'00, Holography 2000, vol. 4149, pp. 56-62, 2000.
15. D. Vila, *Folding Time into Space – Novel Approaches to Integral Holography*, Proceedings of SPIE'93, Practical Holography VII: Imaging and Materials, vol. 1914, Springer, pp. 230-235, 1993.
16. D. Vila, *Holo-Dynamics – Linking Holography to Interactive Media*, Proceedings of SPIE'93, Practical Holography VIII, vol. 2176, Springer, pp. 166-171, 1994.
17. M. Okamoto, H. Ueda, I. Nakamura, E. Shimizu, and T. Kubota, *Multiple Illuminations Method for Rainbow Holograms using an LCD Projector*, Proceedings of SPIE'97, Practical Holography XI, vol. 3011, Springer, pp. 70-75, 1997.
18. K. Yamasakia, M. Okamoto, I. Nakamura, and E. Shimizu, *Fluctuation Hologram with Multiple Images*, Proceedings of SPIE'98, Practical Holography XII, vol. 3293, Springer, pp. 139-144, 1998.
19. J.R. Andrews and W.H. Haas, *Compact Illuminators for Transmission Holograms*, Proceedings of SPIE'89, Practical Holography III, vol. 1051, Springer, pp. 156-159, 1989.
20. J. Bongartz, D. Giel, P. Hering, *Fast 3D topometry in medicine using pulsed holography*, Proceedings of SPIE'02, Practical Holography XVI and Holographic Materials VIII, vol. 4659, 2002.
21. S. Frey, J. Bongartz, D. Giel, A. Thelen, and P. Hering, *Ultrafast holographic technique for 3D in situ documentation of cultural heritage*, Proceedings of SPIE'03, Optical Metrology for Arts and Multimedia, vol. 5146, pp. 194-201, 2003.
22. T. Carlsson, J. Gustafsson, and B. Nilsson, *Development of a 3D camera*, Proceedings of SPIE'99, Practical Holography XIII, vol. 3637, pp. 218-224, 1999.
23. N. Abramson, *Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena*, Applied Optics, vol. 22, no. 2, pp. 215-231, 1983.
24. S.J. Hart and M.N. Dalton, *Display holography for medical tomography*, Proceedings of SPIE'90, Practical Holography IV, vol. 1212, pp. 116-135, 1990.
25. W. Plesniak, M. Halle, S.D. Pieper, W. Wells III, M. Jakab, D.S. Meier, S.A. Benton, C.R.G. Guttman, and R. Kikinis, *Holographic Video Display of Time-Series Volumetric Medical Data*, Proceedings of IEEE Visualization, pp. 78-85, 2003.
26. E.B. Champagne, *Nonparaxial Imaging, Magnification, and Aberration Properties in Holography*, Journal of the Optical Society of America, vol. 57, no. 1, pp. 51-54, 1967.
27. K. Bazargan, *Techniques in Display Holography*, Ph.D. Dissertation, Physics Dept, Imperial College, London University, p. 16, 1986.
28. H. Liu and P. Bhattacharya, *Uncalibrated stereo matching using discrete wavelet transform*, Proceedings of 15th International Conference on Pattern Recognition, vol.1, pp.114-118, 2000.
29. C. W. Slinger, C. D. Cameron, S. J. Coomber, R. J. Miller, D. A. Payne, A. P. Smith, M. A. G. Smith, and M. Stanley, *Recent developments in computer generated holography: towards a practical electroholography system for interactive 3D visualization*, Proceedings of SPIE'04, Practical Holography XVIII, vol. 5290 ,pp. 27-41, 2004.

Enabling View-Dependent Stereoscopic Projection in Real Environments

Oliver Bimber, Gordon Wetzstein, Andreas Emmerling and Christian Nitschke

Bauhaus-University Weimar

Bauhausstraße 11, 99423 Weimar, Germany,

{oliver.bimber, gordon.wetzstein, andreas.emmerling, christian.nitschke}@medien.uni-weimar.de

Abstract

We show how view-dependent image-based and geometric warping, radiometric compensation, and multi-focal projection enable a view-dependent stereoscopic visualization on ordinary (geometrically complex, colored and textured) surfaces within everyday environments. Special display configurations for immersive or semi-immersive AR/VR applications that require permanent and artificial projection canvases might become unnecessary. We demonstrate several ad-hoc visualization examples in a real architectural and museum application context.

1. Introduction and Motivation

Projector-based displays have clearly replaced head-attached displays for most virtual reality (VR) applications. Immersive surround screen displays and semi-immersive wall-like or table-like configurations are being used for visualizing two-dimensional or three-dimensional graphical content.

Today, the majority of augmented reality (AR) applications focuses on mobility. Thus wearable or portable devices have become dominant in this area. However, an increasing trend towards projector-based displays for AR can be noticed. A variety of stationary (e.g., [7,14,18]), movable (e.g., [8,17]) and hand-held (e.g., [20]) projectors have been proposed for displaying graphical information directly on real objects or surfaces instead of performing optical augmentations or video compositions. These approaches, however, mimic real-world situations by applying simplified placeholders of real objects or environments that are optimized for projection (i.e., white surfaces with relatively basic 3D structure, such as multi-planar or parametric surfaces).

State-of-the-art multi-projector approaches are able to display seamlessly blended images onto geometrically and radiometrically complex surfaces [3,10]. Such techniques will open new possibilities for both – augmented reality and virtual reality. The immersive or semi-immersive visualization of stereoscopic three-dimensional graphics will become feasible within real environments and on ordinary surfaces. Special display configurations that require permanent projection canvases or artificial placeholders might become unnecessary.

To achieve this goal, however, several technical problems have to be solved:

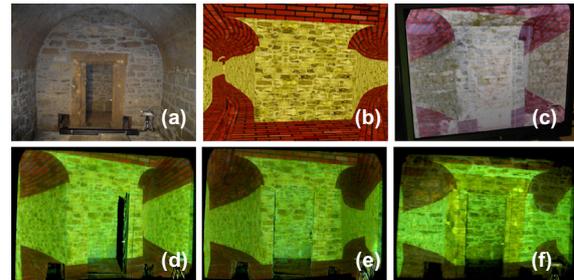


Figure 1: Augmenting water channels in a 9th century water reservoir (a): Screenshot of 3D model showing water pipelines behind wall (b), color compensated projection (c), registered augmented views allowing to see through wall (d,e), color uncompensated image (f).

1. View-dependent geometric correction for projecting images onto geometrically complex surfaces without causing image deformations;
2. Radiometric compensation for projecting images onto arbitrarily colored and textured surfaces to avoid color distortions;
3. Multi-focal projection for displaying images in focus on geometrically complex surfaces;
4. Seamless multi-projector blending for supporting immersive applications and enhancing image quality (e.g., brightness, color, elimination of shadows, etc.).

A correct and consistent depth perception of stereoscopically visualized images on everyday surfaces will fail in many situations if any of these problems cannot be solved adequately. Furthermore we believe that solutions to these problems have to be carried out on a pixel-individual level and in real-time to achieve practical results. In addition, calibration and registration efforts must be kept to a minimum.

In this paper, we want to discuss our first steps towards such an ad-hoc stereoscopic visualization in real environments. Inspired by unstructured Lumigraph rendering, we present an image-based warping method, as well as a refined geometric mapping technique for view-dependent projections onto geometrically and radiometrically complex surfaces. We outline our experiences made with radiometric compensation of stereoscopic projections onto colored and textured surfaces. Additionally, we describe first results of a multi-focal projection method that minimizes blur effects of images displayed on surfaces with varying depth. All these techniques are accomplished at

interactive rates and on a per-pixel basis for multiple interplaying projectors. Finally, we describe application examples and show how all these components can play together to support view-dependent stereoscopic AR/VR visualizations in real environments – without the need of special projection canvases.

We want to present these techniques in the remainder of this paper while discussing the related and previous work on individual topics within the corresponding sections. We do not describe seamless multi-projector blending techniques since these issues have been addressed earlier (e.g., [3, 4]).

2. Radiometric Compensation

If images are projected onto colored and textured surfaces, the projected light is blended with the reflecting surface pigments. This results in a fusion of projected image and underlying display surface texture.

Figure 1 illustrates this problem. Registered images of a three-dimensional scene are projected onto a natural stone wall. Figure 1f shows the blending of the uncorrected images with the underlying surfaces. A fusion of projected stereo-pairs –and consequently a correct depth perception of the displayed scene– is difficult in this situation. The reason for this is that the human visual system strongly relies on the extraction of salient structure features (such as edges, corners, etc.) for estimating disparities. Since for an uncorrected projection the features of the underlying surface can be dominant over features in the displayed images, the depth perception is not consistent. Portions with dominant real features (e.g., the visible real stones and mortar in figure 1f) lead to a perception of the real surface’s depth in these areas. For areas in which image features are dominant, however, the depth of the virtual object can be perceived. Another example is shown in figures 3e-f. Thus a per-pixel radiometric compensation which neutralizes the blending artifacts is essential for visualizing stereoscopic imagery on textured surfaces (cf. figures 1d-e or figures 3g-h).

Recent work on radiometric compensation [3,10,16] uses cameras in combination with projectors for measuring the surface reflectance as well as the contribution of the environment light. These parameters are then used for correcting the projected images in such a way that blending artifacts with the underlying surface are minimized.

Nayar et al. [16], for instance, express the color transform between each camera and projector pixel as pixel-individual 3x3 color mixing matrices. These matrices are estimated from measured camera responses of multiple projected sample images. They can be continuously refined over a closed feedback loop and are used to correct each pixel during runtime. Later, a refined version of this technique was used for controlling the appearance of two- and three-dimensional objects, such as posters, boxes and spheres [10].

Another approach [3] uses single disjoint camera measurements of surface reflectance, environment light

contribution and projector form-factor components for a per-pixel radiometric compensation using hardware accelerated pixel shaders. This method enhances radiometric compensation results by supporting multiple interplaying projectors. Figure 1c illustrates a corrected projector contribution and the observable results are shown in figures 1d-e.

A precise projection of pixels onto corresponding surface pigment is extremely important – especially for noisy surface texture, such as the stone wall in figure 1. Slight misregistrations can cause profound visual color and brightness artifacts, such as bright or dark spots that highlight the real surface features. The human visual system is extremely sensitive to such artifacts, and depth perception of virtual content is disturbed by highlighted real surface features. Note that geometric image distortions which result from such misregistrations are not as critical.

To achieve an adequate registration of projector pixels, we generate a 2D look-up table that maps every pixel from camera space to projector space and vice versa. We apply projected structured light patterns for measuring an unambiguous mapping. All radiometric measurements (surface reflectivity, environment light contribution and projectors’ form factor components) are taken from the perspective of the camera. This ensures a direct look-up in the pixel shader for relating each projector pixel to the parameters of the corresponding surface pigment. Note that this principle is also being applied by related systems [10,16]. This mapping also allows a geometric image correction for the perspective of the camera if images are projected onto non-planar surfaces. In our approach, this is realized with a pixel displacement mapping in a pixel shader.

So far, we have discussed how images can be projected onto geometrically complex, colored and textured surfaces in such a way that they appear correct from the perspective of a single camera (or for observers located within the sweet spot area of the camera). These techniques can easily be extended towards multiple interplaying projectors to enhance the overall image quality [3], cancel out shadows cast by uneven surfaces [3] or carry stereoscopic images. For a single camera calibration, stereoscopic images can be rendered without having information about the viewpoint or the image plane using the method described by Jones et al. [12]. The stereo pairs can be projected with two or more projectors while stereo separation can simply be achieved with LC shutters in front of the projectors’ lenses that are synchronized with active shutter-glasses worn by the users. However, stereo-capable DLP projectors provide high refresh rates (120Hz and more). They support active stereo with a single device and without a loss of brightness that is due to additional LC shutters. For view-dependent applications (e.g., head-tracked stereoscopic visualizations), however, a single sweet spot is not sufficient. Our extension towards a view-dependent correction of geometry and radiometric measurements is described below.

3. Geometric Warping

For projecting multiple undistorted images registered onto planar surfaces the estimation of projector-to-camera homographies is very common (e.g., [6]). Two main techniques exist for geometrically complex surfaces (a good overview can be found in Brown et al. [4]):

1. The per-pixel mapping technique described above is applied if multiple projectors have to be registered precisely for one fixed viewing position [3,10,16].
2. A 3D model of the surface can be acquired (either scanned or manually modeled). All projectors are then registered together with the surface model to a common coordinate system in which the observer is tracked. Projective texture mapping and two-pass rendering can be applied to geometrically warp the view of a moving user into the view of the projectors [19].

As we noted earlier, the first technique provides an adequate precision for a projection onto colored and textured surfaces but does not support arbitrary view points. The second method does support moving users, but suffers from registration errors and is not trivial to realize in practice [4]. Registration errors in the order of 2-3 pixels that lead to geometric distortions on white surfaces can sometimes be tolerated, in contrast to color and intensity artifacts caused on textured surfaces or inter-projector misregistrations. The geometry has to be scanned, and has to be precisely registered to its real counterpart. In addition the projectors have to be calibrated precisely to the surface. Both is difficult to achieve with the precision required by radiometric compensation. Finally, projective texture mapping is based on a simple pinhole camera model and can consequently not handle non-linear effects, such as radial distortion caused by projector lenses.

3.1. Image-Based Approach

To create a correct projection for a single camera, the geometric mapping between camera and projector(s), as well as the radiometric parameters have to be measured during calibration [3]:

Let $P2C$ be a two dimensional look-up table in the resolution of the projector that maps every projector pixel to a corresponding camera sub-pixel (i.e., the image of the projected pixel on the surface). Each $P2C$ map can be computed by reversing the mapping from camera space to projector space ($C2P$). A $C2P$ map has the resolution of the camera and can be measured using structured light range scanning (gray code scanning with phase-shifting in our implementation). Note that $P2C$ and $C2P$ are usually not one-to-one mappings. Thus multiple entries have to be averaged during measuring and inversion – leading to a sub-pixel precise mapping. Furthermore when inverting a $C2P$ map, the resulting $P2C$ map might be incomplete. Thus missing portions have to be interpolated. This is done by triangulation and

off-screen rendering into a 16 bit floating point P-buffer, followed by a read-back into the texture memory.

In addition to the geometric distortion, the radiometric parameters are measured. The surface reflectance and a projector's form-factor contribution are measured by projecting a white image and capturing the camera response (FM). The same is done for measuring the environment light as well as the black-level of each projector by projecting and capturing a black image (EM). Having an original image (O) and a definition of where this image has to appear within the camera view, a pixel shader can perform a pixel displacement mapping using the $P2C$ map in such a way that O appears geometrically undistorted from the perspective of the camera. In addition, every pixel of O can be color corrected by looking up the corresponding radiometric measurements (FM and EM) and performing the necessary computations in the pixel shader. The parameter textures $P2C$, FM , and EM have to be measured for every projector that contributes to the final image. They are used by the pixel shader for computing the projector individual response. More details are provided in [3].

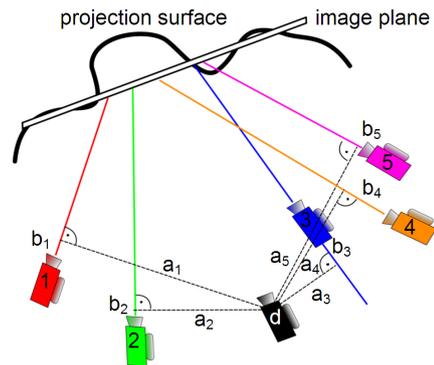


Figure 2: Image-based rendering approach: Five source cameras (1-5), one destination camera (d).

Figure 2 illustrates how this approach can be extended towards view-dependent image-based rendering that was inspired by unstructured Lumigraphs [5]:

We measure the set of parameters for multiple, unstructured source camera positions: $P2C_j$, FM_{ij} , and EM_{ij} , where i is the projector index and j is the camera index. Measuring one set takes about 20 seconds in our implementation. But this depends on the latency of the applied camera which –if no hardware synchronization is provided– has to be soft-synchronized to the structured light projection. The camera is tracked and we store its position and orientation together with each corresponding parameter set. Figure 2 illustrates this for five unstructured source camera positions. For rendering the image correctly, we have to define where the image plane will appear in space. This can be done once before calibration by interactively aligning a 3D model of the image plane at the desired position in the real

environment¹. If the camera is moved, the registered image plane has to be rendered according to the new camera perspective. We offer two different image plane types: An on-axis image plane remains at a fixed position in space but its orientation is updated in such a way that it is perpendicular to the vector that is spanned by the camera's position and the central position of the image plane. The orientation and position of an off-axis image plane remains constant in space – no matter where the camera is located. The example in figures 3e-h show an off-axis image plane situation.

If we assume that we want to render a correct image only for one of the calibrated source cameras j , the following step is performed: The projection of the image plane into the camera's perspective has to be computed first. This is done by off-screen rendering the registered image plane model from the perspective of this camera. The image plane is shaded with texture coordinates that allow a correct perspective mapping of the original image O onto it. These texture coordinates range from $u=0..1$ and $v=0..1$ for addressing and displaying the entire image O . In the following we want to refer to this image plane texture as IP . Projector-individual pixel shaders can then carry out the following tasks: For each pixel of projector i find the corresponding radiometric parameters in FM_{ij} and EM_{ij} using P_i2C_j . Then find the corresponding pixel of the original image O by referencing P_i2C_j first to look-up the texture coordinate of O in IP . Using this texture coordinate, perform a look-up in O . Having all parameters, the color correction is performed [3] and the pixel is displayed.

For a novel destination camera position that does not match any of the source camera positions, however, all parameters have to be computed rather than being measured: The geometric and radiometric parameter textures, as well as the direction vector for this novel camera perspective are interpolated from the measured parameters of the source cameras. A new image plane texture IP is then rendered from this interpolated perspective. For a correct interpolation, the position of the destination camera is projected onto the direction vectors of the source cameras. Two distances can now be computed for each source camera j (cf. figure 2): The distance from the destination camera to its projection points on each source cameras' direction vector (a_j). And the distance from the destination camera's projection points to each source cameras' position (b_j). These distances are used for computing penalty weights for each source camera with:

$$p_j = \alpha a_j + (1 - \alpha) b_j \quad (1)$$

Note that all distances a_j and b_j have to be normalized over all source cameras before computing the penalties. The factor α allows weighting the contribution of each distance. Since a shift of the destination camera along a source direction vector causes less distortion than a shift away from it, a_j must be weighted higher than b_j . We

¹ Visual feedback for this process can be provided by rendering the image plane perspectively correct into the video stream of one source camera.

chose $\alpha = 0.75$ in our implementation. Note that neither the orientation of the destination camera, nor the intrinsic parameters of source or destination cameras have to be taken into account for computing the penalty weights.

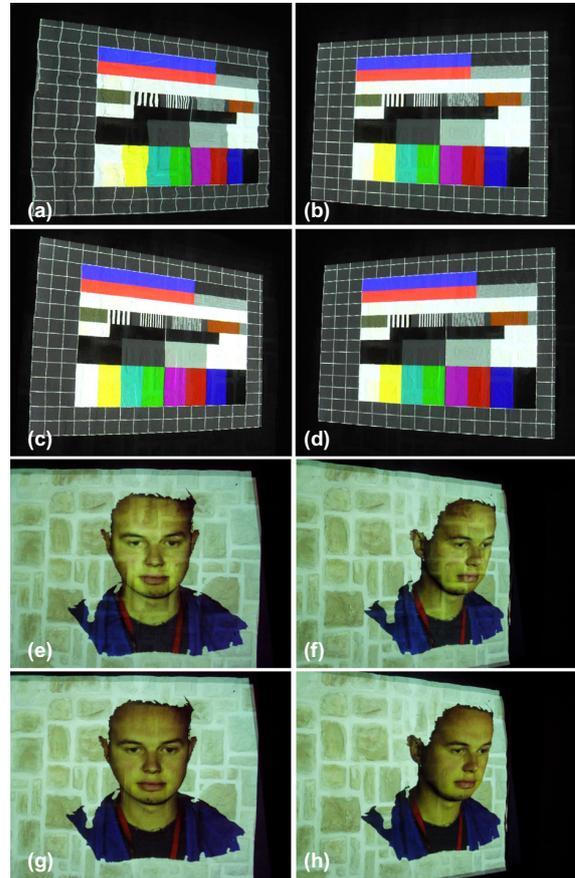


Figure 3: Geometric image warping: Image is distorted (a) when leaving the sweet spot of a single calibration camera (b), but not if a view-dependent correction is applied (c-d).

Geometrically corrected stereo-pairs of 3D scene: without (e-f) and with (g-h) radiometric compensation.

From all source cameras, we select a subset of k cameras with the smallest penalties. Only these k source cameras are considered for sampling the destination camera's new parameter textures. For each of the k source cameras, a weight factor can be computed with:

$$w_j = \left(1 - \frac{p_j}{\max_{pk} p_j} \right) \frac{1}{p_j} \quad (2)$$

where \max_{pk} is the maximum penalty among the k selected source cameras. Note that all weights have to be normalized after being computed. This implies that the source camera with the largest penalty (\max_{pk}) among the k selected ones is weighted with 0. A source camera with close-to-zero penalty is first weighted with a value

approaching infinity, but is then mapped to 1 after normalization.

The parameter textures and interpolated direction vector for the destination camera can now be computed, rather than being measured. This is performed with the pixel shaders by interpolating each parameter entry t_j of P_i2C_j , FM_{ij} , EM_{ij} , and the original direction vectors among the k selected source cameras as follows:

$$t_d = \sum_j^k w_j t_j \quad (3)$$

Note that look-ups in FM_{ij} and EM_{ij} have to be carried out with the original (non-interpolated) P_i2C_j map while look-ups in IP have to be done with the interpolated projector-to-camera map. This allows the computation of the geometric warping, the image plane projection, and the radiometric parameters (surface reflectance, environment contribution and black-level) for a novel destination camera. Note that for completely diffuse surfaces, the radiometric parameters do not change. Weak specular effects, however, are taken into account with this method as well. To handle a flexible number of source cameras, the pixel shaders are not hard-coded, but dynamically generated and loaded onto the graphics card during runtime. This happens only if k is modified.

Figures 3a-b show the image distortion that can be perceived if leaving the single sweet spot area (figure 3b). Figures 3c-d illustrate that these distortions are not present if the above described rendering method is used. Instead of applying an off-axis projection, simple texture mapping of the image on 2D quad on the cameras' image planes has been used in this example. Note that these images are all color corrected before projecting them onto the surface. Five source camera positions have been calibrated for this example.

Besides a static two dimensional image, interactively rendered stereo-pairs of a three-dimensional content can be displayed in O . These images can be rendered on-axis or off-axis for the defined image plane from the tracked position of the destination camera. This is shown in figures 3e-f for a color uncorrected scene and in figures 3g-h for the same scene with color correction enabled. Note that the tracked destination camera is equivalent to the tracked OpenGL camera defined for the perspective of the observer – not a physical camera. In the following we refer to it as the observer camera. Figures 1 and 8 illustrate other examples.

This image-based approach provides a pixel-precise mapping for view-dependent projections onto geometrically and radiometrically non-trivial surfaces. It does not require a 3D model of the surface or a registration of the projectors within a common three-dimensional coordinate system. However, it does correct non-linear projector distortions.

3.2. Geometry-Based Approach

If the geometry of the projection surface is known, a two-pass rendering technique can be applied for projecting the image undistorted [19]: In the first pass, the image that has to be displayed is off-screen rendered

from the perspective of the observer (or the camera). This image O is then read back into the texture memory. In the second step, the geometry model of the display surface is texture-mapped with O while being rendered from the perspective of the projector. For computing the correct texture coordinates that ensure an undistorted view from the perspective of the observer projective texture mapping is applied. Using this hardware accelerated technique allows to dynamically compute a texture matrix that maps the 3D vertices of the surface model into the texture space of the observer's perspective.

It is easy to see that the precision of this method strongly depends on the quality of the surface model and on an adequate registration between surface model and projectors. As mentioned earlier, misregistrations of 2-3 pixels that lead to geometric image distortions can be tolerated. Performing the radiometric compensation with wrong parameters, however, leads to immediately visible color and intensity artifacts. They make the fusion of stereo pairs difficult. Below we describe a variation of this geometric rendering method that is applicable for view-dependent radiometric compensation:

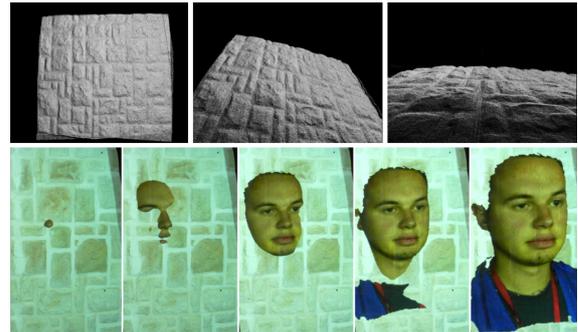


Figure 4: Computed geometry map (top) used for consistent occlusion between real and virtual objects (bottom).

For N camera perspectives, we compute $N P_i2C_j$ maps for each projector. The P_i2C_j maps of any two camera positions lead to a single geometry map estimation that covers only the overlapping surface areas which can be seen from both cameras. A geometry map has the resolution of a projector. It stores 3D positions (in world coordinates) of the corresponding surface points (SP_i) at all projector pixels that project onto these points. The disparities for computing the surface points are given through a definite mapping from a single projector pixel to the two corresponding camera pixels. Consequently, $N^2/2 - N/2$ partial geometry map estimations can be computed for each projector.

Using an unstructured and variable set of source camera positions, we find it hard to acquire a high-quality 3D model of the surface right away if the cameras are not mounted precisely. Small tracking errors (especially in orientation) lead to unusable initial geometry map estimations. We compute a final geometry map composition from all initial geometry map estimations by minimizing the geometric distance

of corresponding 3D points via a least-square optimization. The extrinsic camera orientations are therefore continuously re-adjusted until a global minimum is found.

Doing this for each projector allows us to compute viable projector individual geometry maps (GM_i) that cover all visible portions of the surface (cf. figure 4-top). They are aligned with their real counter part, but don't have to be explicitly registered to the projectors. A definite mapping of 3D surface points to projector pixels is provided implicitly through indexing GM_i . The radiometric parameters (EM_{ij} and FM_{ij}) are measured for each camera-projector combination. The parameter textures that belong to the same projector are then averaged and C_j2P_r -mapped to projector individual look-up textures (EM_i and FM_i) that correspond to the indexing of GM_i .

We compute a texture matrix (TM) that transforms the 3D surface points into the perspective of the observer camera. This matrix is a composition ($TM = N*I*E$) of extrinsic (E : position and orientation transformations) and intrinsic (I : perspective projection) parameters of the observer camera followed by a transformation from normalized device coordinates into normalized texture space ($N=\text{translate}[0.5,0.5,0.0]*\text{scale}[0.5,0.5,1.0]$ for OpenGL). Note that the same matrix is also applied for texturing by conventional projective texture mapping methods. The rendering of the geometry from the perspective of the projector(s), however, is different in our approach:

During runtime a full-screen quad is rendered for each projector to trigger the fragment processing of every projector pixel. The following computations are then performed by the pixel-shader of each projector: For every projector pixel, the corresponding surface point SP_i is looked up in GM_i and is mapped into the perspective of the observer camera with $TM*SP_i$. As illustrated in figure 4-bottom, consistent occlusion effects can be achieved by performing a depth test with the transformed scene points and the depth map of the virtual scene. Being in the camera space, the pixel of the original image O can be referenced in the corresponding look-up textures, as described in section 3.1. Remember that for performing a look-up in O , the texture coordinates of the defined image plane have to be referenced. Thus the image plane texture IP has to be computed for the perspective of the observer camera and passed to the pixel shader exactly as described in section 3.1. The radiometric parameters can be looked-up in EM_i and FM_i . Note that EM_i , FM_i , and GM_i have projector resolution.

In contrast to conventional projective texture mapping approaches, this variation ensures that the look-up of the radiometric parameters for each projector pixel always matches with its corresponding surface pigment. Only the mapping into the observer's camera perspective depends on the quality of the estimated geometry map. This, however, can only cause a geometric misalignment of the image – but no color or intensity artifacts. Intrinsic and extrinsic parameters of the projectors do

not have to be acquired. Non-linear projector distortions are corrected by this method as well.

4. Multi-Focal Projection

Conventional projectors can focus on single focal planes only. Projecting onto complex 3D surfaces with varying depth causes blur effects that wash out image features which are needed for correct disparity-based depth perception.

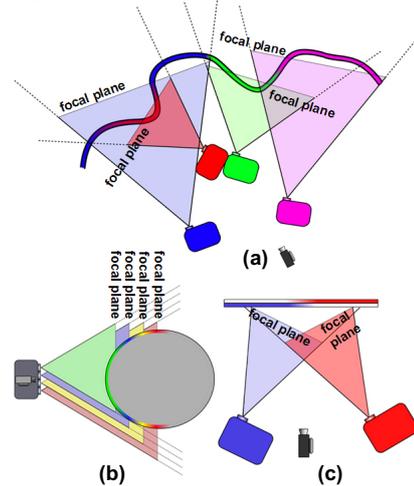


Figure 5: Multi-focal projection concept: Multiple projectors span different focal planes in space for different configurations (a,b,c). The focus value of each projector-pixel is estimated. A final image with minimal focus error is composed from multi-projector contributions.

To overcome this problem, planetariums or some VR displays [2] apply laser projectors instead of conventional projectors for displaying focused images on curved projection screens (e.g., domes or cylinders). Direct-writing-scanning-laser-beam projectors provide a very large focal depth. The cost of a single laser projector, however, can quickly exceed the cost of 500-700 conventional projectors.

Today's off-the-shelf projectors use structured light in combination with an integrated CCD sensor for measuring and adjusting an average focus of the entire image automatically to a planar projection screen. For projecting onto surfaces with varying depths, however, the focus of every single pixel has to be measured and adjusted individually.

Our multi-focal projection concept was inspired by the work on synthetic projection confocal imaging of partially occluded environments [13], and by the work on creating graphical blur effects using superimposed projectors [15]. These approaches follow a reverse idea –creating synthetic defocus effects on real surfaces using projectors– and address different application goals. Akeley et al. [1], for instance, presented a desktop volumetric display prototype that provides multiple focal distances in space. Note that this is not possible with our solution, since we can only focus on the physical display

surface and not on arbitrary virtual planes in space that do not belong to the display surface. A focused stereoscopic projection on non-planar surfaces, however, allows a correct fusion of stereo pairs – but does not solve the general inconsistency between accommodation and convergence that is related to stereoscopic display approaches.

Our basic idea for achieving a multi-focal projection is illustrated in figure 5: Multiple conventional projectors with different focal planes, but overlapping images are used. They can be arbitrarily positioned in the environment (cf. figure 5a), or can be integrated into a single projection unit (cf. figure 5b). The focus error created on the surface has to be estimated for each projector pixel. If this is known, a final image with minimal defocus can be composed from the contribution of all projectors. We want to present our initial results below.

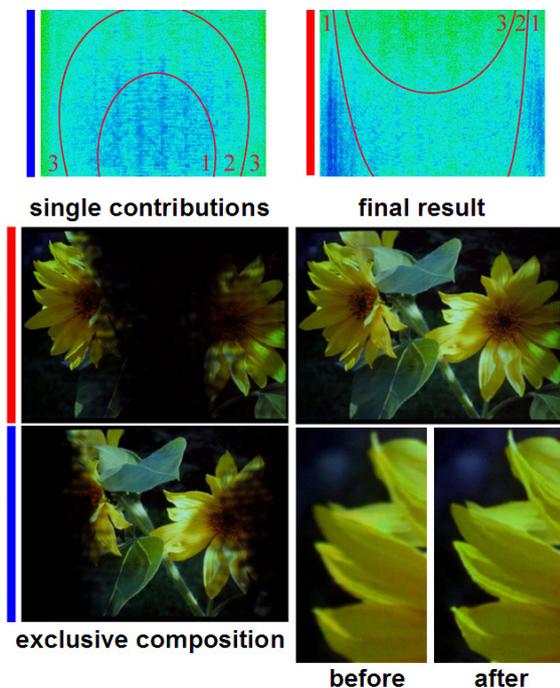


Figure 6: Initial multi-focal projection result: Two projectors (red and blue) project onto a cylindrical surface as illustrated in figure 5b. The blue focal plane intersects the surface in the front; the red focal plane intersects the surface in the middle. The measured focus values of both projectors (top) are used to determine the single pixel contributions that result in a final image with minimal defocus. Color coding: blue to green = best to worst.

4.1. Per-Pixel Focus Estimation

Several multi-focus image reconstruction techniques exist that compose a photograph with a virtually large focal depth from several registered photographs with differently focussed image portions (e.g., [9]). These

algorithms have to estimate the focus values for each pixel using the image content only (e.g., via spatial image gradients). Estimating the focus values for an active display, however, allows simplifying this problem by analyzing displayed structured patterns (e.g., [21]). As mentioned above, such techniques are already being applied by off-the-shelf projectors for providing auto-focus capabilities, and are only capable of estimating an overall focus value for the entire image. Our intention, however, is to estimate an individual focus value for each pixel of each projector.

To determine the per-pixel focus values we project a uniform grid of sample points onto the projection surface in such a way that the points appear geometry and color corrected in the camera view. This is important because the focus estimation has to be independent of the projectors' and camera's distance to the surface, or the surface's color and texture. We achieve this by applying the geometry correction and radiometric compensation techniques described above. The result is a rectified and color corrected image of the projected sample point grid, whereby all points must have a uniform size, color and intensity – unless they are blurred because they are out of focus.

For estimating every sample point's blur we measure its color corrected and normalized point spread within a search window surrounding each point. The search windows have a uniform size and are also geometry corrected for the camera view. Based on the normalized point spread data, we have implemented and evaluated a variety of focus estimators that analyze the spread's area increase, and loss of intensities or high frequencies. The result is a relative focus value for every sample point. If the discrete sample points are then swept over the entire area of the search windows a focus value for every pixel in camera space can be estimated.

Figure 6 shows an example of two projectors with different focal planes creating an image of minimal defocus on a cylindrical surface (cf. figure 5b). The measured focus values (cf. figure 6-top) have been color coded and grouped into three areas (1=best, 2=medium, 3=worst focus) to give a better impression of the focus behavior. Note that the asymmetry along the vertical axis results from the vertical off-axis alignment of the projection frustum. Therefore, the image is always defocused more at the top than at the bottom. The circular shape of the focus areas are due to slight radial focus variations caused by the projector's lens system and due to the cylindrical shape of the surface.

The focus values are mapped into the perspective of the corresponding projector (via the $P2C$ map) and are stored in a texture with projector resolution. We refer to this texture as focus map FOM_i . Note that the constant focal depth of the camera can increase the blur in our measurements. These camera specific focus effects are constant for all projector measurements. However, we want to compute only relative focus values that allow a comparison among multiple projector contributions rather than absolute values. Consequently, constant

camera-specific focus effects do not play a role for our algorithms.

4.2. Image Composition

Having the relative focus values for every projector pixel, a final image with minimal defocus can be composed and displayed from multiple projector contributions in real-time during rendering.

We offer two strategies for image composition: An exclusive (or binary) mapping ensures that a particular surface portion can be covered only by the projector that provides the smallest focus error at this portion. This method is shown in figure 6. The different contributions have to be blended adequately to take the different color responses of the projectors into account. The final composition image appears consistent and portions that were blurred before (i.e., with single projector contributions) are now in focus simultaneously.

Another strategy is to use the normalized focus values for weighting the pixel intensities of each projector. This allows overlapping projections and can consequently provide a higher image quality (e.g., brightness and contrast).

Note that both composition strategies are implemented as projector-individual pixel shaders, and the measured focus maps FOM_i are passed as parameter textures to the shaders. The focus values are view-independent. They are constant for each projector (if projectors and surface are fixed) and do not have to be estimated and interpolated for different perspectives, as it is the case for other surface parameters described in section 3.1.

5. Summary and Discussion

In this paper, we have shown how view-dependent stereoscopic projection can be enabled on ordinary surfaces (geometrically complex, colored and textured) within everyday environments. We pointed out several problems that have to be solved to ensure a consistent disparity-related depth perception and presented several solutions.

As illustrated in figure 7, we want to summarize that a consistent and view-dependent stereoscopic projection onto complex surfaces is enabled by four main components: geometric warping, radiometric compensation, multi-focal projection and multi-projector contributions. Implementing these components as hardware accelerated dynamic pixel shaders allows achieving interactive frame rates.

Radiometric compensation is essential for a consistent disparity-related depth perception of stereoscopic images projected onto colored and textured surfaces. We have outlined our experience made with stereoscopic projections of three-dimensional scenes using techniques that were previously used for correct projections of two-dimensional monoscopic images. Furthermore we have described how these techniques can be extended to support view-dependent rendering, and noted that an

adequate registration precision is as important in this case as for fixed-view rendering.

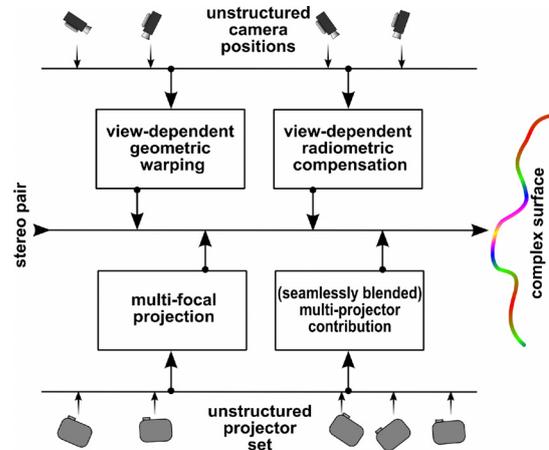


Figure 7: Main components for enabling consistent stereoscopic projection on everyday surfaces.

We presented an image-based warping technique and a refined geometric mapping method that both provide the required precision for radiometrically compensated projections of stereo pairs onto geometrically and radiometrically complex surfaces. Both techniques provide view-dependent rendering at interactive frame rates and support the correction of non-linear projector distortions.

The data that is required for the image-based method is relatively large (for $k=N$ best camera positions: $3N$ textures for each projector plus 2 global textures), but is quick and easy to generate. Another advantage of this method is that geometric and radiometric parameters can be treated in exactly the same way. Such a method can serve as a general rendering framework that will allow computing additional surface parameters (such as view-dependent specular effects) in a uniform manner.

The complexity of this method, however, scales with the number of projectors and the number of camera positions. The data required for the geometry-based approach is relatively small (3 textures per projector plus 2 global textures) but is more difficult to produce. The reason for this is that the quality of the geometry maps depends strongly on the precision of the scanning technology. In case a tracked camera is used instead of a professional range scanner, the precision of the tracking system is a dominant factor. New radiometric parameters are not determined for novel observer positions. Thus view-dependent effects are not taken into account. The advantage of this method, however, is that its complexity scales only with the number of projectors. In addition, the geometry map allows creating consistent occlusion effects.

Note that the number of parameter texture (and consequently k) is limited by today's shader hardware. In general we found that for our experiments $k=4-5$ interpolation cameras (i.e., parameter sets) were sufficient to achieve good results with the image-based

method for large-scale projections (e.g., the examples shown in figures 3 and 8). This also approaches the maximum number of textures today’s pixel shaders can process simultaneously. However, a manifold of source camera positions can be calibrated to cover a large projection space without affecting the rendering performance. Only the k best cameras are dynamically selected and passed to the pixel shaders for interpolation. For computing usable geometry maps, however, much more effort had to be put into. The rendering performances for both methods under the same conditions were comparable (32-16 fps for stereo with two projectors on an off-the-shelf PC with an nVIDIA FX6800 graphics card). A hybrid method that computes the image warping geometrically, but interpolates the radiometric parameters over discrete sample positions for taking view-dependent effects into account might be most efficient.

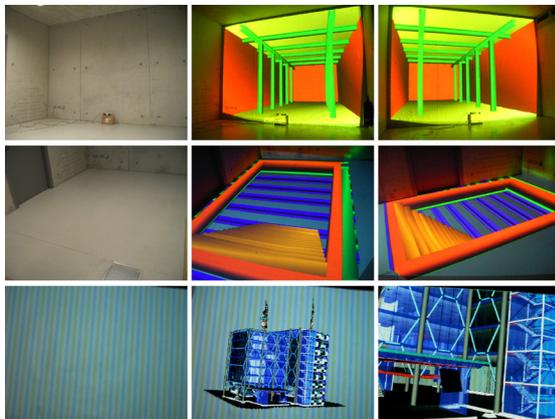


Figure 8: Application examples: Visualizing registered structures of an adjacent room (top) or stairways and lower building level (center) for architectural applications in a realistic environment. Displaying an architectural walk-through on a papered wall in an office (bottom).

We argued that a partially unfocused projection affects disparity-related depth perception and can make the correct fusion of stereo pairs difficult. Since conventional projectors provide only a single focal plane, it is physically impossible to generate sharp images on surfaces with extreme depth variations. We proposed a multi-focal projection method that provides a simple but efficient solution to this problem. Using multiple projectors, different physical focal planes can be generated. In this case, the entire problem can be solved in software – by automatically estimating per-pixel focus values and composing a final image with minimal focus error from multiple projector contributions in real-time. Since next-generation projectors will become very compact (size of a matchbox, using power efficient LEDs with low heat development instead of conventional light-bulbs) and cost effective, an integrated solution of many active stereo-capable DLP projection units into a single device (as outlined in figure 5b) is well imaginable. In this

paper, we have described early results on our multi-focal projection concept. We are currently working on experiments with more complex surfaces using a denser set of focal planes.

To discuss multi-projector techniques (such as chrominance mapping, intensity matching, cross-fading, or shadow removal, etc.) was out of the scope of this paper. These techniques have been described in detail elsewhere. However, we do use these methods. In fact all described techniques are multi-projector capable and are used in multi-projector configurations.

As outlined in the architectural examples shown in figure 8 or the museum-oriented example shown in figure 1, such techniques do not only offer new possibilities for augmented reality and virtual reality. They also allow merging both technologies and give some applications the possibility to benefit from the conceptual overlap of AR and VR. An ad-hoc system configuration that is not constrained to laboratory conditions will increase the practicability of such a tool. The examples shown in figures 1 and 8 have been set up from scratch in a non-laboratory environment in less than one hour (including calibration and hardware installations). Note that the background of the visualized architectural scene in figure 1-bottom has been stenciled out and is intentionally not color corrected. This allows seeing through the 3D mode wherever possible to sense the real environment in addition. The effect is a consistent spatial disparity-based perception of the virtual and the real environment.

However, there are some technical limitations that have to be reported: The fixed resolution of both – cameras and projectors– prevent from measuring and correcting small geometric details and colored pigments that fall below their resolutions. The solution to this problem is to allow a higher spatial resolution (projector and camera) on a smaller surface area. This can be achieved simultaneously by a larger number of stationary projectors and cameras² or interactively by a single tracked projector/camera device. Furthermore off-the-shelf projectors and cameras suffer from a low dynamic range, which makes the capturing and compensation of a large color space impossible. Potential solutions are multi-channel projectors that provide a high-dynamic range in combination with multi-spectral imaging technology [11]. High-dynamic range or dynamic range increase techniques represent further software solutions that can enhance the camera measurements. The high black-level of conventional projectors also makes it difficult to produce absolutely dark areas. For multi-projector configurations the black-level of each projector is being added. This prevents us currently from using a large number of overlapping projectors for creating very bright images with a high contrast that can compensate all possible pigment colors. Optical filters can reduce the black-level – but they will also reduce the brightness. In some situations, local

² Or a single mobile camera covering a larger number of sample positions sequentially.

contrast effects (dark areas surrounded by bright areas are perceived darker than they actually are) reduce this problem on a perceptual level.

Diffuse materials that perfectly absorb light in one or more bands of the spectrum are not well suited for a radiometric compensation approach. Fortunately, such materials are not very common in the real world. Most diffuse surfaces scatter at least a small portion of the light being projected onto them. Thus this challenge reduces to the question of how much light we are able to project for achieving the desired result. Actually, it is the contrary problem whose search for solutions belongs to our future work: the sub-surface scattering of light that needs to be dynamically corrected for creating consistent immersive experiences in everyday environments.

References

- [1] Akeley, K., Watt, S.J., Grishick, A.R., and Banks, M.S., A Stereo Prototype with Multiple Focal Distances, Proc. of ACM Siggraph'04, pp. 804-813, 2004.
- [2] Biehling, W., Deter, C., Dube, S., Hill, B., Helling, S., Isakovic, K., Klose, S., and Schiewe, K., LaserCave – Some Building Blocks for immersive Screens, Proc. of Int. Status Conference on Virtual - and Augmented Reality, Leipzig, 2004.
- [3] Bimber, O., Emmerling, A., and Klemmer, T., Embedded Entertainment with Smart Projectors, IEEE Computer, vol. 38, no. 1, pp. 56-63, 2005.
- [4] Brown, M., Majumder, A., and Yang, R., Camera-Based Calibration Techniques for Seamless Multi-Projector Displays, IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 2, pp. 193-206, 2005.
- [5] Buehler, C., Bosse, M., McMillan, L., Gortler, S.J., and Cohen, M.F., Unstructured Lumigraph Rendering, Proc. of ACM Siggraph'01, pp. 425-432, 2001.
- [6] Chen, H., Sukthankar, R., and Wallace, G., Scalable alignment of large-format multi-projector displays using camera homography trees, Proc. of IEEE Visualization, pp. 339-346, 2002.
- [7] Cotting, D., Naef, M., Gross, M., and Fuchs, H., Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display, Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 100-109, 2004.
- [8] Ehnes, J., Hirota, K., and Hirose, M., Projected Augmentation – Augmented Reality using Rotatable Video Projectors, Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'04), pp. 26-35, 2004.
- [9] Eltoukhy, H.A. and Kavusi, S., A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction, Proc. of SPIE Electronic Imaging, pp. 332-341, 2003.
- [10] Grossberg, M.D., Peri, H., Nayar, S.K., and Bulhumeur, P., Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 1, pp. 452-459, 2004.
- [11] Hill, B., (R)evolution of Color Imaging Systems, Proc. of 1st European Conference on Color in Graphics, Imaging and Vision (CGIV'02), pp. 473-479, 2002.
- [12] Jones, G., Lee, D., Holliman, N., and Ezra, D., Controlling Perceived Depth in Stereoscopic Images, Proc. of SPIE, vol. 4297 Stereoscopic Displays and Virtual Reality Systems VIII, 2001.
- [13] Levoy, M., Chen, B., Vaish, V., Horowitz, M., McDowall, I., and Bolas, M., Synthetic Aperture Confocal Imaging, Proc. of ACM Siggraph'04, pp. 825-834, 2004.
- [14] Low, K-L., Welch, G., Lastra, A., and Fuchs, H., Life-Sized Projector-Based Dioramas, Proc. Symp. Virtual Reality Software and Technology, IEEE CS Press, pp. 93-101, 2001.
- [15] Majumder, A. and Welch, G., Computer Graphics Optique: Optical Superposition of Projected Computer Graphics. Proc. of Eurographics Workshop on Virtual Environment/ Immersive Projection Technology, pp. 209-218, 2001.
- [16] Nayar, S.K., Peri, H., Grossberg, M.D., and Belhumeur, P.N., A Projection System with Radiometric Compensation for Screen Imperfections, Proc. of International Workshop on Projector-Camera Systems, 2003.
- [17] Pinhanez, C., The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces, Proc. Ubiquitous Computing, Springer, pp. 315-331, 2001.
- [18] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H., The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays, Proc. of ACM Siggraph'98, pp. 179-188, 1998.
- [19] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., and Fuchs, H., Multi-projector displays using camera-based registration, Proc. of IEEE Visualization, pp. 161-168, 1999.
- [20] Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., and Forlines, C., iLamps: Geometrically Aware and Self-Configuring Projectors, Proc. ACM Siggraph, ACM Press, pp. 809-818, 2003.
- [21] Tsai, D.M. and Chou, C.C., A fast measure for video display inspection, Machine Vision and Applications, vol. 14, pp. 192-196, 2003.

Compensating Indirect Scattering for Immersive and Semi-Immersive Projection Displays

Oliver Bimber, Anselm Grundhöfer, Thomas Zeidler, Daniel Danch and Pedro Kapakos
Bauhaus-University Weimar

ABSTRACT

We present a real-time reverse radiosity method for compensating indirect scattering effects that occur with immersive and semi-immersive projection displays. It computes a numerical solution directly on the GPU and is implemented with pixel shading and multi-pass rendering which together realizes a Jacobi solver for sparse matrix linear equation systems. Our method is validated and evaluated based on a stereoscopic two-sided wall display. The images appear more brilliant and uniform when compensating the scattering contribution.

CR Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation – Display Algorithms; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Radiosity / Virtual Reality

Additional Keywords: Compensating Indirect Scattering, Interreflections, Projection Displays, Reverse Radiosity, Image Correction, Real-Time Rendering, GPU Programming.

1 INTRODUCTION AND MOTIVATION

Today, a variety of different immersive and semi-immersive projection displays exist to support Virtual Reality (VR) applications. The pallet reaches from single-sided walls, over single- and two-sided workbenches to surround screen displays, such as cylindrical or conical displays [26], spherical domes and multi-sided CAVEs [9].

Beside such dedicated screens, other research activities focus on using projection-optimized [27] or unmodified [3] real surfaces for realizing embedded stereoscopic visualizations in everyday environments.

Several image correction techniques, such as geometric warping, photometric mapping (chrominance and luminance) and edge blending have been developed for multi-projectors displays. While some of them are implemented on dedicated hardware, others benefit from the continuously increasing capabilities of consumer graphics cards. A recent state-of-the-art overview on such techniques was presented by Brown, et al. [5].

Many projection screens are concavely shaped to cover a large portion of the visual field with graphics for giving the user an immersive experience. The light that is projected onto these screens is not only diffused towards the direction of the observer. A fraction of it scatters also to other screen portions – causing a blending with the directly projected image. This *indirect scattering* causes the displayed image to appear partially inconsistent and washed out (cf. figure 1). Specialized screen materials can direct the light towards a particular direction only, and consequently reduce the effects of scattering on other portions. However, such screens do not provide a consistent

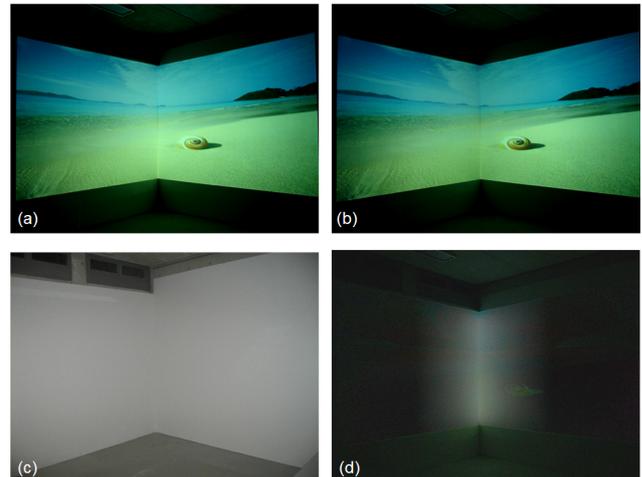


Figure 1. Front-projection onto two-sided wall display (c): Normal projection (a) and projection with compensated indirect scattering (b). The compensated scattering is visualized (d¹) by subtracting photograph (b) from photograph (a).

brightness for arbitrary perspectives.

In this paper, we introduce a method that minimizes these indirect scattering effects in real-time – leading to more brilliant and uniform images.

Our main contribution is the mathematical and algorithmic formulation of a *reverse radiosity* approach for compensating first-level and higher-level indirect scattering effects. We approximate the solution numerically with a Jacobi solver for sparse matrix linear equation systems. It is described how this method can be implemented directly on the GPU via pixel shading and multi-pass rendering. Efficient data structures are presented that allow packing the required information into textures which are processed by pixel shaders. Frame-buffer objects (FBOs) are used for a fast exchange of intermediate iteration results and enable computations with floating point precision. We explain quality and performance optimizations of the rendering algorithm, and discuss simplifications for the correction of stereo-pairs, as well as modifications to support light directing screen materials. Our method is finally validated and evaluated on a stereoscopic two-sided wall display. We provide information on limitations and potentials, as well as data on the performance and the quality of our correction method.

2 RELATED WORK

We want to review two main areas of related work: image correction techniques for projection systems, and hardware accelerated radiosity and inverse radiosity approaches.

Email: {bimber,grundhoefer,zeidler,danch,kapakos}
@uni-weimar.de

¹The resulting image is slightly enhanced in contrast.

2.1 Image Correction for Projection Displays

Many multi-projector rendering techniques exist that aim at creating consistent image geometry, intensity and color among multiple projection units. While geometric image warping (e.g., [1]) is important for curved-screen displays, such as cylinders, cones, and domes, photometric correction (e.g., [21]) and edge-blending (e.g., [28]) is essential for multi-projector configurations, such as tiled walls and tiled curved screens. Several of them benefit from programmable rendering pipelines to achieve real-time performance. Photometric mapping, as an example, has been realized with fragment programs to support color and brightness corrections for tiled wall displays and CAVEs that apply digital projection and Infitec’s multi-band color filtering technology for stereo separation [18]. To discuss all of them in detail is out of the scope of this paper. Instead, we want to refer to a state-of-the-art overview that is presented in [5].

For high-dynamic range visualizations in CAVEs, Dmitiev et al. have developed a customized tone-mapping technique [10] that (in addition to other parameters) applies foveal adaptation [24] relative to the screen portion which the observer is looking at. This is done instead of executing tone-mapping independently for each projector, which would lead to inconsistencies in the corners.

Crosstalk reduction techniques for active [17,20] and passive [16] stereo displays are other correction examples that estimate the ghost images for the left and right views, and subtract them from the opposite views before displaying them. Fixed-function pipeline rendering (e.g., standard OpenGL) is used for achieving interactive frame rates [16]. To support non-linear correction functions in real-time, however, the application of programmable GPUs is obvious, but –to our knowledge– not yet put into practice.

Recent work on radiometric compensation [2,13,22,27] uses cameras in combination with projectors for measuring the surface reflectance as well as the contribution of the environment light. These parameters are then used for correcting the projected images in such a way that blending artifacts with the underlying surface are minimized. This allows using projection screens with imperfections or even arbitrary (i.e., geometric and radiometric complex) surfaces.

Our goal is to compensate *global illumination* effects in the projected images that result from indirect scattering. Previously presented correction techniques, such as the ones summarized above, correct *local illumination* effects that result from inconsistent direct projection. Although they are not immediately related, they can all be carried out as a pre-process of our method.

2.2 Hardware Accelerated Radiosity and Inverse Radiosity

In computer graphics, matrix radiosity is computed by solving a linear equation system with a finite element model [11]. Numerical techniques, such as Jacobi and Gauss-Seidel iteration have been used for this. Throughout the past decades, data structures and rendering algorithms have been optimized and adapted to the evolution of graphics hardware to reach interactive frame rates.

The hemicycle method by Cohnen and Greenberg [7], for instance, was a first example for speeding up visibility and form factor computations using the depth buffer. Kelle, as another example, used hardware-accelerated OpenGL to speed-up radiosity computations without following a finite element approach [14]. He performed a quasi-Monte Carlo particle simulation on the CPU and used local illumination with point light sources at the corresponding particles’ positions.

Recent methods achieve interactive frame rates by implementing the corpus of computations on modern GPUs. Only a few pre-computations are still carried out on the CPU. Jacobi

iteration is particularly well suited for an implementation on GPUs’ streaming-oriented SIMD architectures, but the Gauss-Seidel method requires less iterations. In addition, floating point textures allow encoding all parameters with the required precision. More information on matrix solvers and linear algebra for GPUs is provided in [4,19].

Carr et al., have realized radiosity with a Jacobi solver implemented as a pixel shader [6]. A meshed texture atlas is used for re-sampling and mapping the computed radiosity to the scene.

While Carr et al. pre-compute form factors, Coombe et al. have realized a progressive refinement radiosity algorithm that performs the majority of computations on the GPU – including visibility and form factor computations [8]. While the patch visibility is determined in a shooting-oriented manor with a vertex shader, the form factors as well as the radiosity are computed for the receiving patches in a gathering-oriented way with a pixel shader. An adaptive subdivision is realized on the CPU with texture quad-trees.

Inverse radiosity [29] is an inverse global illumination technique that approaches to recover the diffuse surface reflectance from Lambertian environment patches. It requires that the geometry, the lighting conditions and the radiance distribution are known. Photographs of the environment’s surfaces with light sources, as well as high dynamic range techniques are used to capture the radiance distribution. If this is known, the diffuse albedo of all surface patches can be computed analytically.

In contrast to this, our algorithm compensates the amount of light that is indirectly scattered from other surfaces. It does so by iteratively solving a linear equation system that consciously estimates the scattering under a given direct illumination situation. The subtraction of these estimates from the direct illumination leads to a progressive compensation of the scattering. To avoid confusions with inverse radiosity, we want to refer to this process as *reverse radiosity*.

We assume that the surfaces don’t change over time. Consequently, visibility and form factors can be pre-computed. Pixel shading and multi-pass rendering are applied to implement Jacobi iteration for solving a sparse matrix linear equation system that results in a new *direct illumination*. The main difference to traditional radiosity methods that compute *radiance* (or *radiosity*), or to inverse radiosity techniques that estimate *reflectance* is that we compute the direct illumination that has to be projected to compensate visible scattering.

3 REVERSE RADIOSITY FOR SCATTERING COMPENSATION

As mentioned earlier, concavely shaped projection screens scatter a fraction of light to other screen portions. The amount of indirect illumination adds to the directly projected image. A simple two-patch front-projection example is illustrated in figure 2 (this applies also to back-projection):

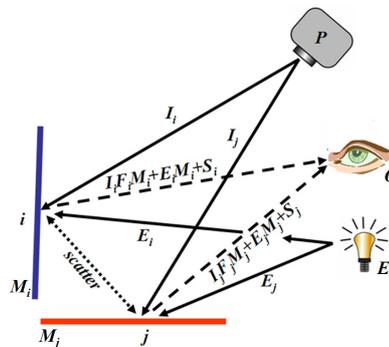


Figure 2. Scattering of light between two surface patches.

Assuming the projector to be a point light source, the energy fraction F of light I that leaves the projector and arrives at a screen patch (i or j) depends on the geometric relation between the projector and the surface. A simple representation of what is known as *projector-to-surface form factor* can be used for approximating this fraction: $F=f*dA*cos(\alpha)/(r^2*\pi)$ where α is the angular correlation between the light beam and the patch normal, dA is the patch's differential area, and r the distance (square distance attenuation) between the light source and the patch surface. The scaling factor f allows an additional fine-tuning to cope with unknown projector properties (e.g., exact intensities).

The direct illumination is reflected depending on the material reflectance M of the screen patches. In addition to the reflected environment light EM , the reflected indirect scattering S contributes to the final image that is perceived by an observer (O). Note, that in contrast to the direct illumination I , the environment light contribution E cannot be controlled. It contains stray light coming from the surrounding room (i.e., in case of semi-immersive projection displays, such as two-sided workbenches) as well as the black-level of the projector.

Problematic is the fact that the amount of scattering depends on the projected image I and vice versa. As explained earlier, the reverse situation (i.e., the global radiosity of each surface element is determined based on a known illumination I) is computed by solving the well known radiosity equation system² (either on the CPU or on the GPU):

$$B_i = E_i + M_i \sum_{j,j \neq i}^p B_j F_{ji} \quad (1)$$

where B_i is the radiosity and M_i the reflectance of patch i , and F_{ji} denotes the *patch-to-patch form factor* that describes the amount of radiosity which is transferred from patch j to patch i .

Below, we describe a *gather-oriented reverse radiosity* method which modifies I in such a way that indirect scattering effects are compensated. We want to assume projection screens with Lambertian reflectance. The form factors and the visibility can be computed off-line.

3.1 First-Level Scattering

We subdivide the entire screen surface into P discrete uniform patches (details on the data structure are presented in section 3.5). If the screen geometry is known, the patch-to-patch form factor can be computed with $F_{ji}=f*dA_i*cos(\alpha_i)*cos(\alpha_j)/(r_{ji}^2*\pi)$, where α_i and α_j are the angles between the ray connecting the two patches' centers and their normals, r_{ji} is the patch-to-patch distance, and dA_i is the differential area of patch i . The scaling factor f allows an additional fine-tuning again to cope with unknown screen properties, and to encode patch-to-patch visibility ($f=0$ if invisible).

The amount of radiance that scatters indirectly to patch i as a result from directly illuminated patches j can be described as (with respect to figure 2):

$$S_i = \sum_{j,j \neq i}^p (I_j F_j M_j + E_j M_j) F_{ji} M_i \quad (2)$$

We want to refer to this as the *first scatter level*. Physically, however, light is scattered back and forth between the patches. How to take these higher scatter levels into account is described in section 3.2.

To compensate *first-level scattering*, the following computation can be performed:

$$I_i^{l+1} = \frac{1}{F_i M_i} (R - E_i M_i - S_i^l) \quad (3)$$

where R is the original image to be displayed.

As mentioned above, the difficulty in this case is that the projected image I depends on the amount of scattering S and vice versa. A simple analytical solution to equation 3 does not exist. Rather than that, this linear equation system can be solved numerically by approximating I through several (l) iterations:

```

compute  $I^0$  //direct illumination
 $l=0$  //number of passes
repeat
  compute  $S^l$  (for  $I^l$ ) //eqn. 2
  compute  $I^{l+1}$  with  $S^l$  //eqn. 3
   $l=l+1$ 
until exit condition is fulfilled
display  $I^{l+1}$ 

```

Algorithm 1. Compensating first level scattering.

Initially, I^0 is computed for the direct case. With $I^0=(R-EM)/FM$ the environment light and projector's form factor contribution (i.e., the non-linear luminance distribution from the image center to its edges) can be compensated. Thereby, R is the original image to be displayed. If this is not necessary, $I^0=R$ assumes a constant luminance distribution and a dark environment.

The result I^0 is used in the first iteration to compute a first estimate of scattering, and to compensate the environment light and the approximated indirect illumination (equation 3). The result I^1 of this first iteration is then used in the second iteration, and so on. This process is repeated until I converges (i.e., the results of two consecutive iterations do not reveal significant differences) or the solving process exceeds a predefined maximum time slot. Details on appropriate exit conditions are explained in section 3.3.

A numerical solution to equation 3 can be computed directly on the GPU via pixel shading and multi-pass rendering that together realizes a Jacobi solver for sparse matrix linear equation systems. The shaders compute equations 2 and 3 in each pass and update I successively. After the exit condition is fulfilled, the final image is displayed. This is summarized in algorithm 1 (pixel shaders are highlighted in blue).

To efficiently exchange the intermediate results of different passes, I is rendered off-screen into an FBO that is then bound to the shaders of the following pass. This avoids time consuming read-back operations and enables computations with floating point precision.

Note that E , M , and F are constant. They do not directly appear in algorithm 1, but have to be considered when applying equations 2 and 3. They are passed as parameter textures to the corresponding shaders.

3.2 Higher-Level Scattering

Equation 4 presents an example that extends equation 2 to the first two scatter levels for a patch i (i.e., the radiance that is scattered directly from the patches j to patch i , and the radiance that is scattered from all patches k over all patches j to patch i):

$$S_i = \sum_{j,j \neq i}^p (I_j F_j M_j + E_j M_j) F_{ji} M_i + \sum_{j,j \neq i,k,k \neq j}^p [(I_k F_k M_k + E_k M_k) F_{kj} M_j] F_{ji} M_i + \dots \quad (4)$$

In the general case, this can be rewritten into a recursive form:

$$S_i^h = \sum_{j,j \neq i}^p S_j^{h-1} F_{ji} M_i \quad (5)$$

² Gathering-oriented, and re-formulated in our notation.

Note, that S_i^h indicates the h -th scatter level at patch i (i.e., the reflected radiance at patch i that can be contributed to scattering which arrives over $h-1$ intermediate patches). The first scatter level (S_i^1) is defined in equation 2. With respect to equation 3 the scattering of the first H levels can be compensated with:

$$I_i^{l+1} = \frac{1}{F_i M_i} \left(R - E_i M_i - \sum_{h=1}^H S_i^h \right) \quad (6)$$

Thereby, I^0 is the initial direct illumination as explained in section 3.1.

Since equation 5 shows that higher scatter levels can be derived from lower scatter levels, all of them can be computed sequentially for the same direct illumination I^l within each render pass l . Having all scatter levels computed, I^{l+1} can be updated according to equation 6. Finally, if the exit condition is fulfilled, the result is displayed. This is summarized in algorithm 2.

```

compute  $I^0$  //direct illumination
 $l=0$  //number of passes
repeat
  compute  $S^{l,l}$  (for  $I^l$ ) // eqn. 2
  for  $h=2$  to  $H$  //scatter level
    compute  $S^{h,l}$  from  $S^{h-1,l}$  // eqn. 5
  endfor
  compute  $I^{l+1}$  with  $S^{l..H,l}$  // eqn. 6
   $l=l+1$  //number of passes
until exit condition is fulfilled
display  $I^{l+1}$ 

```

Algorithm 2. Compensating first H scatter levels.

As for the implementation of algorithm 1, pixel shaders (highlighted in blue) perform the main operations for each patch, and FBOs are used to exchange intermediate textures (I and S) to avoid texture read-backs.

3.3 Exit Conditions

We have implemented two complementary exit conditions for algorithms 1 and 2:

The first exit condition terminates the solver if the images resulting from two consecutive passes do not differ by more than a predefined threshold (e.g., by more than one gray scale unit). Such an exit condition ensures a *constant quality* of scattering compensation over time and can be efficiently implemented on the GPU by using *occlusion queries*: A pixel shader can compare the difference between the two consecutive frames and discard pixels whose corresponding r,g,b differences lie below a predefined *pixel-discrepancy threshold*. All other pixels are written into an FBO³. After this, the query result delivers the number of diverging pixels that have been written into the FBO. This number is compared with a predefined *image-discrepancy threshold* which leads to the decision to exit the solver and display the result or to continue solving.

The second exit condition terminates the solver if the solving process exceeds an assigned time slot. This ensures a *constant solving rate* for scattering compensation. One possibility is to measure the amount of time that is required for the first solving pass initially. Since, in theory, this is constant for all passes, the maximum number of passes that fit into the provided time slot can be determined. However, due to potential hardware-dependent irregularities, such as caching operations, the time required for

each pass might differ slightly. An alternative is to measure the time for each pass individually and subtract it from the time slot until it is exhausted.

Both exit conditions can be combined. Thereby the solver is determined if either the image quality does not change significantly *or* the time slot has been exhausted.

3.4 Interactive Rendering

It is clear that algorithm 1 or algorithm 2 cannot be solved in real-time if the patch resolution P equals the pixel resolution of the original image to be projected. Consequently, all textures (E, I, S, M, L, R , and F_i) have to be down-sampled to a resolution that can be processed at interactive rates. The final projection image can then be computed as follows:

$$I_Q^{l+1} = I_Q^0 - \uparrow (\downarrow I_Q^0 - I_P^{l+1}) \quad (7)$$

Thereby, I_Q^0 is the initial direct illumination image in the original image resolution (Q), I_P^{l+1} is the solved image that compensates scattering in patch resolution P , and \uparrow and \downarrow indicate image operators for up- and down-sampling (from Q to P and vice versa) with bi-linear or bi-cubic filtering. The result I_Q^{l+1} that contains the compensation of the direct illumination as well as of the scattering is finally being displayed.

Note that simply up-sampling I_P^{l+1} into the original image resolution and displaying it directly would lead to a visible patch structure in the projected image. These artifacts are avoided by subtracting the up-sampled scattering portion $\uparrow (\downarrow I_Q^0 - I_P^{l+1})$ with the direct illumination in the original resolution (I_Q^0).

Note also that only patches with patch-to-patch form factors larger than a predefined threshold are selected. This leads to additional resources and performance optimizations and does not affect the outcome's quality significantly. Thus, the patch-to-patch form factor matrix F_{ji} is a *sparse matrix* and not completely filled. This is explained in more detail below.

For stereoscopic rendering, it is sufficient to blend the stereo-pairs into a single image that is used by a single solving step, instead of solving the left and the right stereo image individually. Due to the small difference caused by the parallax shift in the down-sampled image, the variation of both solutions is neglectably small. A single solving step, however, increases the overall rendering performance. The solved result is up-sampled and applied to both stereo images, as explained above. An appropriate blending function is $(left+right)/2$.

3.5 Optimized Data Structures

The down-sampled textures are in the patch resolution P . They are two-dimensional because 2D textures are optimized for pixel shaders. Their corresponding values at each pixel can be addressed with the same patch index u,v . While F_i contains single projector-to-patch form factor values, the other textures store r,g,b values.

Especially the patch-to-patch form factor matrix F_{ji} can become as large as P^2 . This does not only lead to performance issues of the above algorithms, but also to storage problems. As mentioned earlier, F_{ji} is a sparse matrix, and the number and distribution of entries depends on a pre-selected *patch-to-patch form factor threshold*. Only patches that scatter enough energy to another patch are considered. All other entries are left empty. To avoid storing the empty entries, the remaining patch-to-patch form factors are compressed into a single texture by concatenating them as illustrated in figure 3.

³ Rendered off-screen (not displayed).

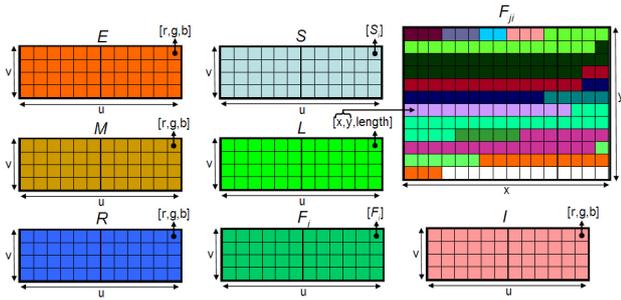


Figure 3. Data structures: The horizontal and vertical patch resolutions can be arbitrarily chosen and are parameterized over the screen surface. A look-up texture maps from u,v space to the x,y space of a texture that stores the patch-to-patch form factors.

A look-up texture L allows mapping them back to individual patches. Thus, to find form factor components of all patches that scatter enough light to patch u,v the entry $L(u,v)$ gives the x,y coordinate of the first scattering patch as well as the number of following scattering patches (*length*). The texture F_{ji} is composed automatically during an offline procedure. Once computed, its last row can contain $X-1$ empty entries in the worst case, if X is F_{ji} 's horizontal resolution. The horizontal texture resolution cannot exceed the maximum supported texture size (currently $X=4096$ for Nvidia or $X=2048$ for ATI). However, it should be chosen in such a way that the number of empty entries is minimized.

Note that all data structures and data values are relative to the entire screen surface. Thus the screen surface has to be parameterized over u and v to become compatible with our matrix radiosity approach. This is simple for parametric surfaces, such as spherical or cylindrical displays. For multi-plane displays, such as two-sided workbenches, or CAVEs, the individual projection screens have to be parameterized altogether within the same u,v space. For our two-sided wall display, as an example, one screen surface can be parameterized with $[u=0..0.5, v=0..1]$, and the other one with $[u=0..0.5, v=0..1]$. Further screen surfaces can be encoded accordingly.

The final image is displayed by texture-mapping it onto a geometric representation of the screen surface, and rendering this scene from the perspective of each projector. If multiple projectors are overlapping, proper cross-fading between the images has to be ensured. This requires an intrinsic and extrinsic registration of all projectors in 3D space. For simple surfaces, such as planes, the final image can simply be texture mapped onto an image grid and interpolated in screen space. The registration can be conducted in 2D space by manually selecting key points (i.e., grid points) directly in the screen space of each projector and assigning them to the corresponding texture coordinates manually. Note that these are common registration techniques for projection displays. While image warping is sometimes realized directly in the hardware of professional projectors or with external devices, it can also be carried out by the graphics card.

For performance reasons, we use 16bit floating point textures and FBOs instead of a slower, but more precise 32bit depth.

3.6 Light-Directing Projection Screens

Some projection screens are designed to radiate more of the projected light towards their normal direction rather than the same amount in all directions. Holographic projection screens, actively shuttered projection screens with phase dispersed liquid crystal layer (as used for the blue-c display [12]), or BARCO's diffuse and bright (DAB) screens are examples.

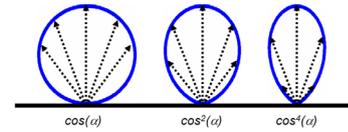


Figure 4. Angular dependency of form factors for regular projection screens (left), and exponentiated cosine function for approximating diffuse light directing projection screens (center and right).

The behavior of such screens can be approximated by our algorithm with a simple exponentiation of the cosine functions when computing the patch-to-patch and projector-to-patch form factors (cf. figure 4). The normal vector of the patches can also be tilted to simulate an alternation of the radiation's principle direction.

4 EVALUATION

In this section, we want to validate our method and analyze its performance and quality parameters. The evaluation has been carried out under the following conditions:

The two-sided wall-display (figure 1c) serves as a testbed for our experiments. Two InFocus DepthQ active-stereo DLP projectors with a maximum resolution of 800x600 pixels each are applied to display stereo-pairs on each of the 3mx2m walls. A 2.8Ghz Pentium IV with 1GB RAM and a Nvidia FX3400-Quattro (PCI-Express with 256MB on-board memory) drive the projectors. For head-tracked applications, we apply an Origin DynaSight infrared tracking system.

We assume that the projection screens are perfectly white and have a Lambertian reflectance. We also ignore the projectors' black-level and consider the surrounding environment to be completely dark. Consequently, we use $M=1$, $E=0$, and all form-factors are computed with a cosine exponent of 1.

4.1 Validation of Method

To validate our method, and to show its potentials as well as its limitations, we have projected a light pattern consisting of the primary colors (red, green and blue) and their complements (magenta, yellow and cyan) onto a corner section of the walls (cf. figure 5a).

A high-resolution digital camera was used to capture a photograph of the reflected light before (U) and after (C) correction. These two photographs have been taken under exactly the same conditions (i.e., from the same perspective and under the same camera settings). When subtracting C from U the amount of correction can be visualized (cf. figure 5b).

It is worth discussing the effects that appear in figure 5b in a bit more detail:

As shown in figure 5a, the primaries are projected onto the right wall while their complements are displayed on the left wall. The projected background surrounding the color patterns is white.

We want to differentiate between the directly projected light D , and the light S that is scattered. For our interpretation of figure 5b, we want to consider only colors (not intensities) and just the first scatter level.

Scattering can be compensated efficiently on screen portions whose direct illumination D is equal or larger than the indirect illumination S scattered onto them in all color channels.

If this is not the case, the compensation of scattering will lead to clipping in one or more channels.

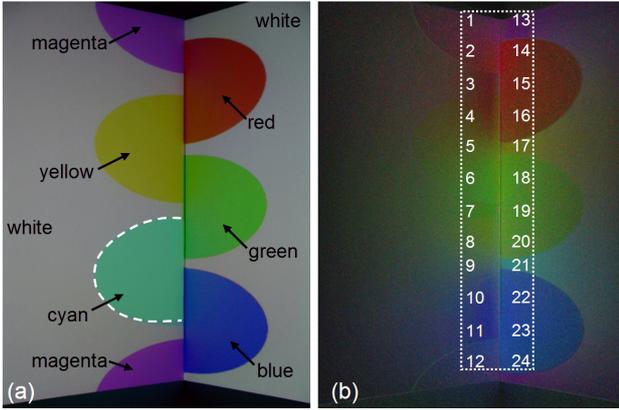


Figure 5. Projected primaries and complements (a). Subtracting a photograph of the corrected projection from a photograph of the uncorrected projection results in figure 5b'.

Figure 5b has been divided into 24 areas (Π). The parameters D and S , as well as the result of the correction ($U-C$) within each area are outlined in table 1. Let us discuss an example:

The direct illumination in areas 14, 15, and 16 is red ($r,g,b=1,0,0$). On the opposite side of this color pattern we have a direct illumination of magenta ($1,0,1$) in area 2, white ($1,1,1$) in area 3, and yellow ($1,1,0$) in area 4. The scattering onto areas 2-4 is red. Since the direct illuminations of areas 2-4 are all larger or equal to the amount of scattered light, the scattering can be completely compensated. The result in $U-C$ must consequently be the scattered light only. This is obviously the case, since the red bleeding in areas 2-4 is well visible. The scattered light can be fully compensated in all areas on the left side, except in areas 1, 5, and 9 (due to the larger white scattering from the right side). If we consider the opposite situation, area 2 scatters magenta, area 3 scatters white, and area 4 scatters yellow onto the red areas (14-16). In all these three cases, the compensation of the scattering is clipped since only the red channel can be further reduced. Consequently, $U-C$ will reveal only a difference in the red color channel, which is also well noticeable in figure 5b. This is also the case for the other two primaries on the right side.

Table 1 presents the effects for all areas and highlights the clipping situations on both sides in bold. The framed area in table 1 corresponds to the framed area in figure 5b. It can be noticed that our theoretical thoughts are reflected in reality.

Note that the example in figure 5 is certainly an extreme one. In general the displayed scenery causes much less clipping. In many cases the colors on opposite sides are similar or even equal, and scattering can be removed efficiently (cf. figures 1 and 7, for example).

It is important to mention that compared to the uncorrected image, the visual quality of the corrected image will never degrade. In extreme situations (e.g., two different primaries bleeding onto each other), a scattering cannot be removed completely in all color channels. In such cases the corrected image does not differ much from the uncorrected one.

4.2 Performance

Figure 6 illustrates the solving performance of our method with a disabled performance-based exit condition. It was measured on the hardware described above, and by rendering the scene shown in figure 7 for the two-sided wall display. The scene is rendered four times for stereo projection, (two stereo images for two image planes). The native image resolution is 1600x600 pixels (800x600 pixels for each projector in a horizontal alignment).

Π	S	D	$U-C$	$U-C$	D	S	Π
1	1,1,1	1,0,1	1,0,1	1,0,1	1,1,1	1,0,1	13
2	1,0,0	1,0,1	1,0,0	1,0,0	1,0,0	1,0,1	14
3	1,0,0	1,1,1	1,0,0	1,0,0	1,0,0	1,1,1	15
4	1,0,0	1,1,0	1,0,0	1,0,0	1,0,0	1,1,0	16
5	1,1,1	1,1,0	1,1,0	1,1,0	1,1,1	1,1,0	17
6	0,1,0	1,1,0	0,1,0	0,1,0	0,1,0	1,1,0	18
7	0,1,0	1,1,1	0,1,0	0,1,0	0,1,0	1,1,1	19
8	0,1,0	0,1,1	0,1,0	0,1,0	0,1,0	0,1,1	20
9	1,1,1	0,1,1	0,1,1	0,1,1	1,1,1	0,1,1	21
10	0,0,1	0,1,1	0,0,1	0,0,1	0,0,1	0,1,1	22
11	0,0,1	1,1,1	0,0,1	0,0,1	0,0,1	1,1,1	23
12	0,0,1	1,0,1	0,0,1	0,0,1	0,0,1	1,0,1	24

Table 1. Analysis of effects shown in figure 5 within the 24 areas (Π): The direct illumination (D) and the first-level scattering (S) are indicated in r,g,b triplets. The effects in the framed area of figure 5b are shown in the framed section ($U-C$). The highlighted values indicate clipping. Only colors (not intensities) are considered.

The blue lines are the measurements for the monoscopic mode – rendering and solving one image per frame. The green lines are the measurements for the stereoscopic mode, rendering left and right stereo images, solving one image containing the blended stereo pair, and applying the result to both stereo images. The different line types illustrate the performance for varying patch resolutions.

The first entry is the frame rate (y-axis) achieved without scattering compensation. The time required to render the scene without compensation was 100fps (mono) and 55fps (stereo). The scatter level is increased over the x-axis. It can be observed that the frame rate drops with an increasing patch resolution. A patch resolution of 64x64 patches leads to a maximum solving performance of around 20fps (for both –mono and stereo). Lower resolutions, such as 32x32 and 16x16 patches led to maximal frame rates of 48fps/35fps and 63fps/42fps (mono/stereo).

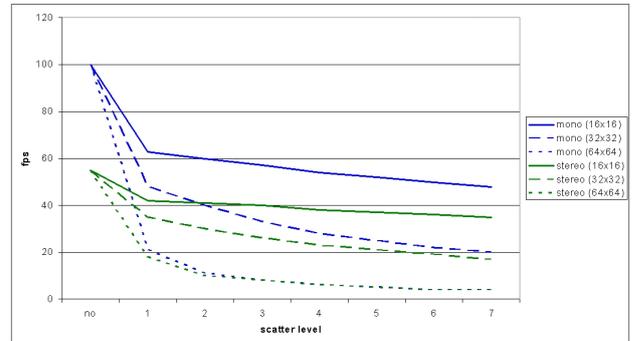


Figure 6. Solving performance measurements for monoscopic and stereoscopic two-sided projection and different patch resolutions.

For a resolution of 16x16 patches, the loss of performance was between 23.6% (1 scatter level) and 36.3% (7 scatter levels) for our example running in the compensated stereo mode (compared to the uncompensated stereo mode). For a patch resolution of 32x32, the performance loss was between 36.3% and 69% for the same mode. As it will be explained in section 4.3, a correction of scattering above level 2 does not lead to significant improvements. Note that although the performance of our method is independent of the scene complexity, the time for rendering the scene has been included in our measurements.

4.3 Quality

It is interesting to make some quantitative statements on the amount of compensated scattering. This can be done by comparing the histogram values of the difference photographs ($U-C$) over different scatter levels with the histogram values of the photograph showing the uncorrected projection.

If we do this for the examples shown in figures 1 or 7, the average amount of compensated first-level scattering (estimated over the entire image region) is approximately 9% (luminance) relative to the uncorrected projection. Note, that the scattering is regionally different. It is larger at screen intersections, or at the screens' lines or points of symmetry. Thus, approximately 28% of scattering has been compensated for the 50cm zone surrounding the screen intersection in figure 1.

If the second-level scattering is considered as well, we achieve an additional compensation of ~1% (estimated again over the entire image region). If we then perform a correction of up to level 7, we receive a further total compensation of only ~0.5% (i.e., from level 2 to level 7). This indicates that a compensation of scatter levels larger than 2 does not lead to large qualitative enhancements that would justify the corresponding loss of performance.

If the patch resolution is sufficiently high (e.g., $\gg 16 \times 16$ in our experiments) a further increase will not change the amount of scattering compensation significantly. Instead, it will influence the precision of our method. In figure 5b, for example, a small color seam is visible directly on the left side of the screen edge. This is due to a too coarse resolution of 16×16 patches. When increasing the patch resolution it disappears. Note that the color seam is amplified in the difference photograph, and that it is barely detectable in the corrected projection.

5 SUMMARY, CONCLUSION AND FUTURE WORK

Indirectly scattered light that adds to directly projected light regionally washes out the displayed images. We have presented a method for compensating indirect scattering effects that occur with concavely shaped immersive or semi-immersive projection displays. The images appear more brilliant and uniform when compensating the scattering contribution (cf. figures 1 and 7). The human visual system is very sensitive to even small variations in brightness. Note, that these slight differences cannot be reproduced well with the low dynamic range photographs presented in this paper, but quantitative statements have been made in section 4.3.

An active stereo application running on a two-sided wall display at 55fps without correction drops to 40fps when our correction method is turned on. The correction, however, compensates ~10% (luminance) first- and second-level indirect scattering over the entire image area and ~30% at the 50cm zone close to the screen intersection. We believe that this is an

acceptable tradeoff – especially when considering further accelerations enabled by upcoming graphics hardware. A correction of scatter levels larger than 2 did not result in significant improvements in our experiments.

The amount of compensated scattering depends on the display surface and on the presented content. A corrected image will not degrade in quality compared to the uncorrected image. If only little scattering can be compensated due to a potentially dark content, a difference between a corrected and an uncorrected case might not be detectable. By analyzing the image content first, we can extend our algorithm to automatically turn the correction off in such situations. This allows gaining additional performance resources for scene rendering.

Compared to a uniform patch resolution, an adaptive subdivision will increase the performance and enhance the precision of our method. An extension of our algorithm and data-structures towards adaptive subdivision lies ahead of us.

For displays that are driven by multiple rendering nodes, our algorithm could be distributed and solved by multiple GPUs in parallel. A pre-analysis of the surface structure allows determining which portions actually influence each other, and which don't. This makes a clustering of interacting patches possible and leads to multiple independent correction tasks that can be carried out by different rendering nodes.

Our experiments are representative for two-sided screen displays, such as two-sided walls or L-shaped workbenches. To evaluate our method for other display types, such as multi-sided CAVEs, cylindrical/conical, and spherical displays, as well as for geometrically complex surfaces belongs to our future work. Figure 8 illustrates a simulation of our method for such systems.

For a corrected projection onto geometrically and radiometrically complex display surfaces (e.g., textured everyday surfaces) [3], the surface parameters –including the patch-to-patch light transfer– have to be measured via structured light projection and camera feedback. Sen, et al. [25] presented a method that measures the entire 6D light transport through a scene by using a projector and an array of cameras. Reducing this problem to view-independent diffuse surfaces the measurements of reflectance, patch-to-patch form factors, and projector-to-patch form factors should be feasible from a single camera position and within an acceptable amount of time.

Finally, the reduction of scattering at sending patches will also be a challenging extension of our current method – which compensates scattering at receiving patches. This, however, will partially change the intensity and color of the original image. A good balance between the compensation of gathered light and the reduction of scattered light has to be found.

Several VR applications intend to simulate and visualize realistic real world lighting appearance –such as illumination engineering, interior design, or virtual prototyping– with

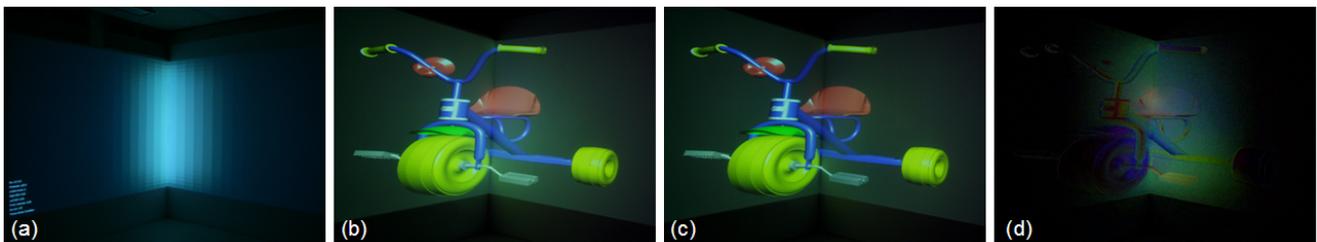


Figure 7. Computed, summed and intensity amplified form factor patches projected onto a two-sided wall display (a). Uncorrected (b) and corrected (c) three-dimensional object. The object does not contain primary colors and the background is light gray. Approximately 9% (luminance) scattering can be compensated. Difference photograph (b-c) shows the amount of correction (contrast amplified by 70%).

immersive or semi-immersive projection displays (e.g., [15]). Much effort is spent on rendering nearly photorealistic global and local illumination effects that arise in the rendered scene. Physical illumination effects that occur on the display's surfaces, however, have also to be taken into account for creating an ultimate lifelike experience.

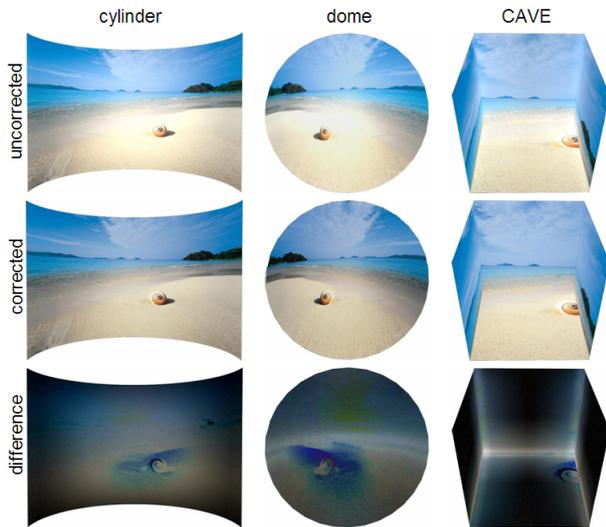


Figure 8. Simulated reverse radiosity for cylindrical display (left), dome (center), and four-sided CAVE (right). The difference images at the bottom are contrast enhanced by 70%.

REFERERNC

- [1] J. van Baar, T. Willwacher, S. Rao, and R. Raskar, Seamless Multi-Projector Display on Curved Screens, Eurographics Workshop on Virtual Environments (EGVE'03), pp. 281-286, 2003.
- [2] O. Bimber, A. Emmerling, and T. Klemmer, Embedded Entertainment with Smart Projectors, IEEE Computer, pp. 56-63, vol. 38, no. 1, 2005.
- [3] O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke, Enabling View-Dependent Stereoscopic Projection in Real Environments. In proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), 2005.
- [4] J.I. Bolz, I. Farmer, E. Grinspun, and P. Schröder, Sparse Matrix Solvers on the GPU: Conjugate Gradients and Multigrid. In proc. of ACM Siggraph'03, pp.917-924, 2003.
- [5] M. Brown, A. Majumder, and R. Yang, Camera-Based Calibration Techniques for Seamless Multi-Projector Displays, IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 2, pp. 193-206, 2005.
- [6] N.A. Carr, J.D. Hall, and J.C. Hart, GPU Algorithms for Radiosity and Subsurface Scattering, In proc. of Siggraph/Eurographics Workshop on Graphics Hardware'03, pp. 51-59, 2003.
- [7] M.F. Cohen and D.P. Greenberg, The hemi-cube: A radiosity solution for complex environments. In proc. of ACM Siggraph'85, pp. 31-41, 1985.
- [8] G. Coombe, M.J. Harris, and A. Lastra, Radiosity on Graphics Hardware, In proc. of ACM Graphics Interface (GI'04), vol. 62, pp. 161-168, 2004.
- [9] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti, Surround-screen projection-based virtual reality: The design and implementation of the cave. In proc. of ACM Siggraph'93, pp. 135-142, 1993.
- [10] K. Dmitriev, T. Annen, G. Krawczyk, K. Myszkowski, and H.P. Seidel, A CAVE system for interactive modeling of global illumination in car interior. In proc. of ACM Symposium on Virtual Reality Software and Technology (VRST'04), pp. 137-145, 2004.
- [11] C.M. Goral, K.E. Torrance, D.P. Greenberg, and B. Battaile, Modelling the interaction of light between diffuse surfaces. In proc. of ACM Siggraph'84, pp. 213-222, 1984.
- [12] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. Vande Moere, and O. Staadt, blue-c: A Spatially Immersive Display and 3D Video Portal for Telepresence. In proc. of ACM Siggraph'03, pp. 819-827, 2003.
- [13] M.D. Grossberg, H. Peri, S.K. Nayar, and P. Bulhumeur, Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System, In proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), vol. 1, pp. 452-459, 2004.
- [14] A. Keller, Instance radiosity. In proc. of ACM Siggraph'97, pp. 49-56, 1997.
- [15] R. Klein, J. Meseth, G. Müller, R. Sarlette, M. Guthe, and A. Balazs, RealReflect - Real-time Visualization of Complex Reflectance Behaviour in Virtual Prototyping, In proc. of Eurographics'03, Eurographics Industrial and Project Presentations, 2003.
- [16] S. Klimenko, P. Frolov, L. Nikitina, and I. Nikitin, Crosstalk Reduction in Passive Stereo-Projection Systems. In proc. of Eurographics'03, pp.235-240, 2003.
- [17] J., Konrad, B. Lacotte, and E. Dubois, Cancellation of image crosstalk in time-sequential displays of stereoscopic video, IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 897-908, 2000.
- [18] W. Kresse, D. Reiners, and C. Knöpfle, Color Consistency for Digital Multi-Projector Stereo Display Systems: The HEyeWall and The Digital CAVE, In proc. of 7th International Immersive Projection Technologies Workshop / 9th Eurographics Workshop on Virtual Environments (IPT/EGVE'03), pp. 271-335, 2003.
- [19] J. Krüger and R. Westermann, Linear Algebra Operators for GPU Implementation of Numerical Algorithms. In proc. of ACM Siggraph'03, pp. 908-916, 2003.
- [20] J.S. Lipscomb and W.L. Wooten, Reducing crosstalk between stereoscopic views. In proc. of SPIE'95, Stereoscopic Displays and Virtual Reality Systems II, vol. 2409, pp. 31-40, 1995.
- [21] A. Majumder and R. Stevens, Perceptual Photometric Seamlessness in Tiled Projection-Based Displays, ACM Transactions on Graphics, vol. 24, no. 1, pp. 111-134, 2005.
- [22] S.K. Nayar, H. Peri, M.D. Grossberg, and P.N. Belhumeur, A Projection System with Radiometric Compensation for Screen Imperfections, Proc. of International Workshop on Projector-Camera Systems, 2003.
- [23] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, The office of the future: A unified approach to image-based modelling and spatially immersive displays. In proc. of ACM Siggraph'98, pp. 179-188, 1998.
- [24] A. Scheel, M. Stamminger, and H.P. Seidel, Tone reproduction for interactive walkthroughs. In proc. of Eurographics'00, pp. 301-312, 2000.
- [25] P. Sen, B. Chen, G. Garg, S. Marschner, M. Horowitz, M. Levoy, and H.P.A. Lensch, Dual Photography, In proc. of ACM Siggraph'05, pp. 745-755, 2005.
- [26] A. Simon and m. Göbel, The i-Cone - A Panoramic Display System for Virtual Environments, In proc. of the 10th Pacific Conference on Computer Graphics and Applications (PG'02), p.3, 2002.
- [27] D. Wang, I. Sato, T. Okabe, and Y. Sato, Radiometric Compensation in a Projector-Camera System Based on the Properties of Human Vision System, In proc. of IEEE International Workshop on Projector-Camera Systems (ProCams'05), 2005.
- [28] R. Yang, D. Gotz, J. Hensley, and H. Towles, PixelFlex: A Reconfigurable Multi-Projector Display System, In proc. of IEEE Visualization (IEEE Vis'01), pp. 68-75, 2001.
- [29] Y. Yu, P. Debevec, J. Maik, and T. Hawkins, Inverse global illumination: recovering reflectance models of real scenes from photographs. In proc. of ACM Siggraph'99, pp. 215-224, 1999.

Multifocal Projection: A Multiprojector Technique for Increasing Focal Depth

Oliver Bimber, *Member, IEEE*, and Andreas Emmerling

Abstract—In this paper, we describe a novel multifocal projection concept that applies conventional video projectors and camera feedback. Multiple projectors with differently adjusted focal planes, but overlapping image areas are used. They can be either differently positioned in the environment or can be integrated into a single projection unit. The defocus created on an arbitrary surface is estimated automatically for each projector pixel. If this is known, a final image with minimal defocus can be composed in real-time from individual pixel contributions of all projectors. Our technique is independent of the surfaces' geometry, color and texture, the environment light, as well as of the projectors' position, orientation, luminance, and chrominance.

Index Terms—Computing methodologies, computer graphics, picture/image generation, digitizing and scanning, display algorithms, image processing and computer vision, digitization and image capture, radiometry, reflectance, sampling, scanning, enhancement, sharpening and deblurring, scene analysis, color, shape.

1 INTRODUCTION AND MOTIVATION

BECAUSE of their increasing capabilities and declining cost, video projectors are becoming widespread and established presentation tools. The ability to generate images that are larger than the actual display device virtually anywhere is an interesting feature for many applications that cannot be provided by desktop screens.

Several research groups exploit this potential by using projectors in unconventional ways to develop new and innovative information displays that go beyond simple screen presentations. Projectors are used to display images on nontrivial surfaces that are geometrically and radiometrically complex. Efficient soft and hardware techniques have been developed to enable:

- geometric warping of images projected onto non-planar surfaces,
- image registration for tiled-screen configurations,
- photometric correction and intensity blending for luminance and chrominance mapping of multiple projectors, and
- radiometric compensation for projecting onto colored and textured surfaces.

Some of these techniques support real-time rendering rates and pixel-precise correction. They allow a correct projection of images onto surfaces, such as cylindrical or dome-like projection theatres, or even everyday surfaces.

Basically, all of them correct colors, intensities and positions of pixels to achieve an overall consistent image. However, one important pixel property remains unaddressed so far: its *focus*.

Today's consumer projectors are designed and engineered to focus images on planar display surfaces. The Schleimpflug principle describes how to offset the focal plane by an off-axis

configuration of the optical system. However, plane-focused images are partially blurred if projected onto surfaces with substantial depth differences. Special lenses, such as f-theta lenses, allow generating focused images on spherical surfaces. Planetariums and some cylindrical projection displays [1] use laser projectors to overcome this problem. Direct-writing-scanning-laser-beam projectors scan almost parallel beams of laser light onto the projection screen. Thereby, the laser beams remain constant in diameter over a large depth range. This results in a remarkable focal depth and in the possibility to display sharp images in sizable dome-like or cylindrical theatres. The cost of a single laser projector, however, can currently exceed the cost of several hundred conventional projectors. The development of laser-diodes is promising and can overcome these drawbacks in future.

Cathode ray tube (CRT) is another emissive projection technology that allows controlling the diameter of the e-beam in terms of adapting their focus to a particular projection screen. An example is Barco's beam spot nullifying technology. Analog CRT projectors have frequently been used for large curved screen projections. However, there is a continuous trend towards the application of transmissive (e.g., LCD) or reflective (e.g., DLP or LCOS) light-valve projectors. They do not allow controlling per-pixel focus parameters of the projected image electronically or optically. Consequently, our multifocal projection technique addresses exactly these types of digital projectors.

If the projection surface is multiplanar, multiple projectors can be arranged in such a way that they focus on individual planar sections (e.g., [15]). This, however, is inefficient and sometimes impossible if the surface becomes geometrically more complex.

In this paper, we want to introduce a novel *multifocal projection* technique that allows to project images with minimal defocus onto geometrically and radiometrically complex surfaces. This is essential to enable, for instance, stereoscopic projections supporting disparity-based depth perception on arbitrary everyday surfaces [3]. Our approach is a pure software solution and is implemented with multiple conventional video projectors (cf. Fig. 1).

• The authors are with the Bauhaus-University Weimar, Bauhausstr. 11, 99423 Weimar, Germany. E-mail: {bimber, emmerling}@uni-weimar.de.

Manuscript received 4 Sept. 2005; revised 23 Dec. 2005; accepted 6 Feb. 2006; published online 10 May 2006.

For information on obtaining reprints of this article, please send e-mail to: tcvg@computer.org, and reference IEEECS Log Number TVCG-0121-0905.

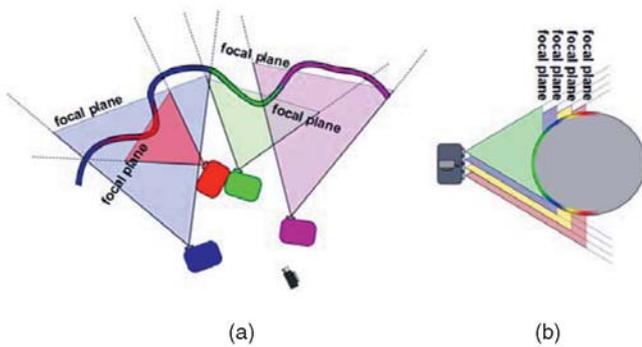


Fig. 1. (a) Multifocal projection concept: unstructured projector set and (b) multiple projection units integrated into a single device.

Multiple projectors with differently adjusted focal planes, but overlapping image areas are used. They can be either arbitrarily positioned in the environment or can be integrated into a single device. The defocus created on a surface is estimated automatically for each projector pixel. If this is known, a final image with minimal defocus can be composed in real-time from individual pixel contributions of all projectors. Our technique is independent of the surfaces' geometry, color and texture, the environment light as well as of the projectors' position, orientation, luminance, and chrominance.

1.1 Outline of the Paper

The remainder of the paper is organized as follows: Section 2 discusses related work that is relevant for the proposed concept. Section 3 reviews our implementation of geometric correction and radiometric compensation that represents the basis for the following techniques. Section 4 describes how geometric and radiometric correction is used to determine normalized intensity spreads of projected sample points on arbitrary surfaces. These intensity spreads are used in Section 5 as input for estimating relative focus values for all projector contributions on each surface portion. Having these relative focus values allows composing a final image with minimal defocus from multiple projector contributions on the GPU. Composition techniques are explained in Section 6. Finally, Section 7 summarizes and discusses our work and gives indications for future directions.

2 RELATED AND PREVIOUS WORK

In this chapter, we discuss related work on multiprojector techniques, focus estimation methods, and radiometric compensation approaches. These techniques are relevant components for our multifocal projection approach.

2.1 Multiprojector Displays

Tiled screen displays apply multiple adjacent projectors for creating large images with a high resolution. Since the images are usually displayed on planar surfaces, projectors can be geometrically registered by estimating projector-to-camera or projector-to-projector homographies. Overlapping image areas are cross-faded, and varying luminance and chrominance parameters are matched to achieve the desired consistency [4]. Cameras are usually applied to measure the corresponding correction parameters. In addition, shadows cast by objects or users can be removed if the image overlap of

multiple projectors is sufficiently large [11]. Projecting images onto nonplanar surfaces requires a different way of geometric warping. If the surface geometry and the projectors' intrinsic and extrinsic parameters are known, two-pass rendering and projective texture mapping can be applied to create geometrically consistent images from arbitrary perspectives [22]. Multiple projectors are applied again to avoid shadow regions. If the projection surface contains high frequencies in its geometry and reflectance, an image-based warping from uncalibrated projectors becomes more accurate than projective texture mapping from calibrated ones [3]. Radial lens distortions can also be corrected by such methods, which is not possible by assuming a simple pin-hole camera model—as it is done with projective texture mapping.

Our multifocal projection concept was inspired by the work on synthetic aperture confocal imaging of partially occluded environments [14] and by the work on creating graphical blur effects using superimposed projectors [16]. These approaches follow a reverse idea—creating synthetic focus/blur effects on real surfaces using multiple projectors—and address different application goals.

2.2 Focus Estimation

Passive and active autofocus techniques are implemented in modern cameras, as well as in projectors with integrated CCD sensors. In contrast to passive techniques, active approaches apply a structured light projection to support more reliable focus estimations on unstructured surfaces. An average focus value is usually estimated for the entire image or a small image region. This estimate is then minimized by mechanically modifying the optics which moves the focal plane forth and back along the optical axis of the system. Since this has to be done fast, focus operators that can be efficiently realized in hardware, such as fast Fourier transformation (FFT), are normally applied.

Digital autofocus techniques determine the blur at edges. The assumption is that while sharp edges represent step functions, blurred edges are (Gaussian) low-pass filtered step functions. Knowing their edge spread functions (ESF),¹ blurred areas can be focused digitally by determining and applying a high-pass reconstruction filter kernel derived from the parameters of the ESF [5]. Some techniques estimate point spread functions (PSF) from the ESF and derive the filter kernels from the PSF rather than from the ESF [12], [13].

General image fusion techniques estimate focus values for every single pixel of multiple sample images taken from the same perspective, but under different focus settings. Eltoukhy and Kavusi [6], for instance, utilize the gradients of a pixel computed within a predefined search window. A final image with maximal sharpness can then be composed from the sample images by selecting the pixels with largest gradients among the corresponding ones. Other approaches decompose the sample images using a wavelet transformation [10]. The resulting image is then being composed from the largest wavelet coefficients and reconstructed via an inverse wavelet transformation.

Similar to the image fusion techniques described above, shape-from-focus methods analyze discrete focus values in differently focused images. The intention, however, is to estimate the depth of the captured surface. While passive techniques [8] rely on the surface texture, active methods apply structured light to enhance the focus estimation. If the

1. Equivalent to a point spread function in one direction.

light pattern is projected through the same optical system as the images are recorded with [20], [18], a geometric correction is not necessary. The focus values can then be derived by applying a Laplace operator to recorded images illuminated with an optimized checkboard pattern.

A multiparameter camera system—similar to the one used in [18], but without projector—has recently been applied for video matting based on defocus [17]. The cameras share an optical center, but vary in focus, exposure, and aperture. The PSF is estimated for deriving defocus values that allow separating the foreground from the background.

In the case of our multifocal projection approach, the projectors and the camera can be placed at arbitrary positions and do not make use of the same optical system. Consequently, the focus estimation must be independent from the intrinsic and extrinsic parameters of camera and projectors, the environment light, as well as from the surface’s geometry and reflectance.

2.3 Radiometric Compensation

Recent work on radiometric compensation uses cameras in combination with projectors for measuring the surface reflectance as well as the contribution of the environment light. These parameters are then used for correcting the projected images in such a way that blending artifacts with the underlying surface are minimized.

Nayar et al. [19], for instance, express the color transform between each camera and projector pixel as pixel-individual 3×3 color mixing matrices. These matrices are estimated from measured camera responses of multiple projected sample images. They can be continuously refined over a closed feedback loop and are used to correct each pixel during runtime. Later, a refined version of this technique was used for controlling the appearance of two and three-dimensional objects, such as posters, boxes, and spheres [9]. Wang et al. [24] adapts this method to the properties of the human vision system by compressing the contrast of the input images. Fujii et al. [7] applies a variation of the closed feedback loop method to handle dynamic environments by applying a coaxial projector-camera alignment.

3 GEOMETRIC CORRECTION AND RADIOMETRIC COMPENSATION

This section reviews our implementation of geometric and radiometric image correction [2]. In contrast to [19], [9], [24], [7] it uses single disjoint camera measurements of surface reflectance, environment light contribution and projector form-factor components for a per-pixel radiometric compensation using hardware accelerated pixel shaders. This method enhances related radiometric compensation methods by supporting multiple interplaying projectors.

In our current implementation, these image correction techniques represent the basis for the multifocal projection approach described in Sections 4, 5, and 6. However, we believe that related methods can be used instead, if they can be extended to support multiple projectors.

3.1 Single Projector

In its simplest configuration, an image is displayed from a single projector (P) in such a way that it appears correct (color, intensity, and geometry) for a single camera view

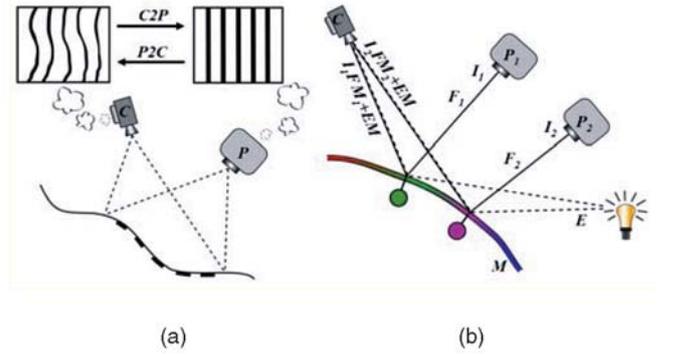


Fig. 2. (a) Geometric calibration and (b) radiometric compensation.

(C). Therefore, the display surfaces must be Lambertian, but can have an arbitrary color, texture, and shape.

The first step is to determine the geometric relations of camera pixels and projector pixels over the display surface. We use well-known structured light techniques (gray code scanning with phase shift in our implementation) for measuring the 1 -to- n mapping of camera pixels to projector pixels (cf. Fig. 2a). This mapping is stored in a 2D look-up-texture having a resolution of the camera, which we refer to as $C2P$ map. A corresponding texture that maps every projector pixel to one or many camera pixels can be computed by reversing the $C2P$ map. We call this texture $P2C$ map. It has the resolution of the projector. The 1 -to- n relations (note that n can also become 0 during the reversion process) are finally removed from both maps through averaging and interpolation.

Once the geometric relations are known, the radiometric parameters are measured. We assume that a light ray with intensity I is projected onto a surface patch with reflectance M . The fraction of light that arrives at the patch depends on the geometric relationship between the light source and the surface. The common projector-to-surface form factor representation is the simplest way to approximate this fraction: $F = f^* dA * \cos(\alpha) / (r^2 * \pi)$, where α is the angle between the light ray and the surface normal, dA is the surface patch’s differential area, and r is the distance between the light source and the surface. The constant scaling factor f allows an additional fine-tuning to cope with unknown projector parameters, such as exact intensities.

Together with the environment light E , the projected light fraction I is blended with the patch’s reflectance M (cf. Fig. 2b):

$$R = EM + IFM. \quad (1)$$

Thus, R is the diffuse radiance that can be captured by the camera. If R , F , M , and E are known, I can be computed with:

$$I = (R - EM) / FM. \quad (2)$$

In our approach, E , F , and M cannot be determined independently. We measure FM by projecting a white image ($I = 1$) and turning off the entire environment light ($E = 0$), and EM by projecting a black image ($I = 0$) under environment light. Note that EM also contains the black-level of the projector.

Since this holds for every discrete camera pixel, R , E , FM , and EM are stored in two-dimensional textures and (2) can be computed² in real-time by a pixel shader on the GPU.

During runtime, pixel displacement mapping is realized by rendering a full-screen quad into the frame buffer of the projector. This triggers fragment processing of every projector pixel. The shader maps all pixels from the projector perspective into the camera perspective (via texture look-ups in the $P2C$ map) to ensure a geometric consistency for the camera view. All computations are then performed in camera space. The projection of the resulting image I onto the surface leads to a geometry and color corrected image that approximates the desired image R for the target perspective of the camera.

3.2 Multiple Projectors

The simultaneous contribution of multiple projectors increases the total light intensity that arrives at the surface. This can overcome the limitations of (2) for extreme situations (e.g., small FM or large EM values) and can consequently avoid an early clipping of I . If N projectors are used, (1) extends to (cf. Fig. 2b):

$$R = EM + \sum_i^N I_i FM_i. \quad (3)$$

Our strategy is to balance the projected intensities equally among all projectors i which leads to:

$$I_i = \frac{(R - EM)}{\sum_j^N FM_j}. \quad (4)$$

This equation can also be solved in real-time by projector-individual pixel shaders (based on individual parameter textures FM_i , $C2P_i$, and $P2C_i$ —but striving for the same final result R). Note that EM also contains the accumulated black-level of all projectors.

4 MEASURING INTENSITY SPREADS

Our goal is to estimate relative focus values caused by each projector on all portions of an arbitrary display surface. Having this information, a final image with minimal defocus can be composed from multiple projector contributions.

This chapter describes our approach of measuring the intensity spreads of projected sample points. They are proportional to the defocus of the sampling projector on the corresponding surface portion.

Like for the geometric and radiometric image correction (Section 3), we apply a structured light projection in combination with camera feedback to estimate the focus values (cf. Fig. 3). However, instead of displaying horizontal and vertical scan lines, a uniform grid pattern of circular sample points is used. Applying points instead of strips also allows determining the defocus parameters in multiple directions simultaneously.

4.1 Precorrection

Displaying a sample point from the projector's view would lead to a color and geometry distorted image of it in the camera's view. This situation would make it impossible to

2. For each color channel separately.

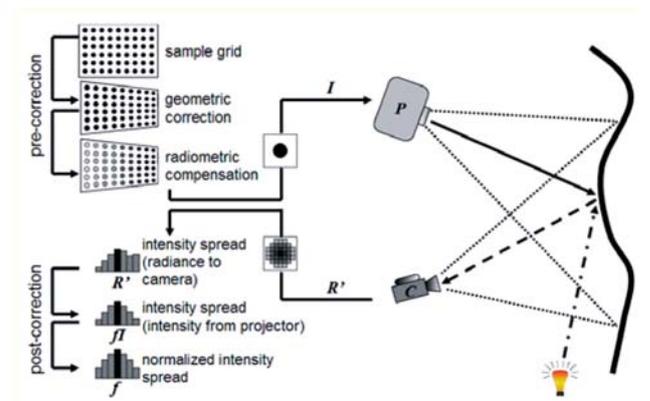


Fig. 3. Measuring intensity spreads on geometrically and radiometrically complex surfaces.

estimate focus values for the corresponding surface area. The reason for this is that it is not detectable in this case whether the recorded intensity spreads of the sample point can be attributed to defocus or to external factors (e.g., geometric or radiometric distortion, blending with the environment light on the surface, or effects related to the projector's position or orientation).

To overcome this problem, we sample the surface from the perspective of the camera instead of from the perspective of the projector. Measuring relative focus values within the same space and under the same conditions enables a qualitative comparison between the contributions of individual projectors. Thus, a sample point is initially defined in the camera perspective and is then geometrically warped into the perspective of the projector via the beforehand determined $C2P$ map (Section 3). In addition, the sample point is color corrected with (2) to compensate for the surface reflectance, the environment light, and the projector's form factor contribution. We set $R > 0$ for pixels that belong to the sample point and $R = 0$ for all other pixels.

If we assume a perfectly sharp image (perfectly focused by the projector and by the camera) the image recorded by the camera contains the initially defined and undistorted sample point. In this case, it has retained its original circular shape of the defined size and appears at a uniform intensity that approximates the defined intensity and color R .

However, due to blur effects (caused either by the projector or by the camera) the shape and the intensity of the sample point are no longer uniform. The resulting intensity spread and the intensity loss can be measured in the camera image. They are proportional to the relative focus of the projector at this point. Measuring the defocus caused by different projectors at the same sample point enables a qualitative and relative comparison and finally an optimal image composition. The camera's parameters (i.e., intrinsic, extrinsic, response, etc.) do not influence the outcome. But, they must not be changed during the measurements for multiple projectors that need to be compared.

4.2 Postcorrection

As explained above, a defocused sample point creates an intensity spread on the surface that is captured in the camera image. The intensities in the blurred area are not projected in a controlled way. They are also blended with the underlying surface reflectance. To estimate the focus

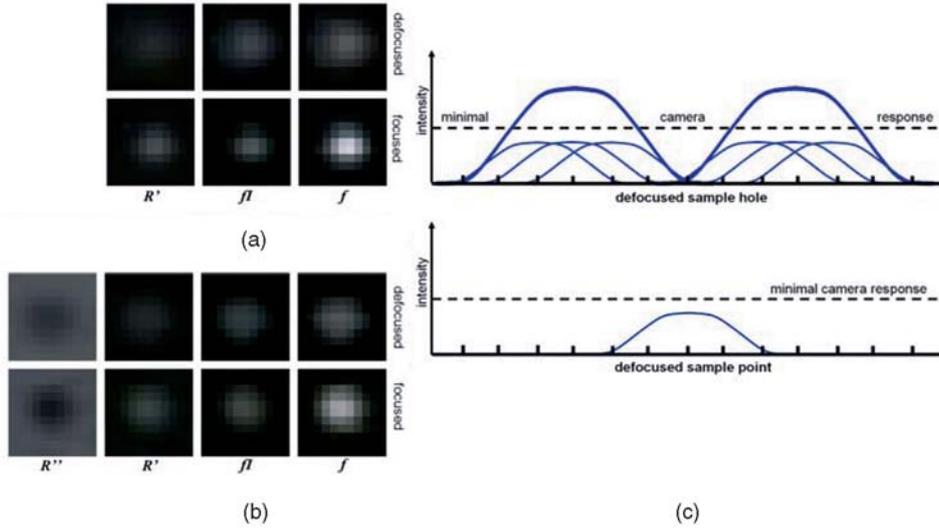


Fig. 4. Intensity spreads during postcorrection: (a) regular sample points and (b) inverse sample holes—defocused and focused case. (c) Intensity accumulation at sample holes.

values of a sample point consistently, the intensity spread has to be normalized in such a way that it becomes independent of the surface reflectance, the environment light and the projectors' form factor contributions at the spread areas. Note that this normalization cannot be carried out during the precorrection step because the intensity spread has to be measured first.

The precorrection applies (2) to project the corrected sample point in such a way that it appears at coordinate x, y in the camera image:

$$I_{x,y} = (R_{x,y} - EM_{x,y})/FM_{x,y}. \quad (5)$$

As outlined earlier, we assume that $R_{x,y} > 0$ for pixels belonging to the sample point, and that $R_{x,y} = 0$ for all other pixels.

The intensity spread $R'_{x',y'}$ in the blurred area x', y' is measured by the camera as well. It results from defocus and is a fraction f of the original sample point's intensity that is blended with the surface reflectance in the blur area:

$$R'_{x',y'} = fI_{x,y}FM_{x',y'} + EM_{x',y'}. \quad (6)$$

To normalize the intensity spread, we have to determine f . For this, we estimate the intensity which the projector would have to produce to create the same reflected radiance $R'_{x',y'}$ with a direct illumination:

$$fI_{x,y} = \frac{(R'_{x',y'} - EM_{x',y'})}{FM_{x',y'}}. \quad (7)$$

Finally, f can be computed by comparing the results of (5) and (7):

$$f = \frac{fI_{x,y}}{I_{x,y}}. \quad (8)$$

If we normalize the pixels in the captured camera image, we will receive a normalized intensity distribution of the sample point (cf. Fig. 4). This is independent of the surface reflectance and of the projector's form factor. Due to limitations of the projectors (brightness and dynamic range) and the camera (response and noise) the maxima of the intensity spreads might not equal one. A second

normalization that lifts their maxima intensities to one³ enhances the quality of focus operators that do not analyze the intensity loss of the spread (this will be explained in more detail in Section 5).

Note that it is not possible to correct the geometry of the blur area for more complicated surfaces. We could transform it into the perspective of the projector using the $C2P$ map. However, this would cause an even larger geometric distortion which is useless for comparing relative focus estimations of different projectors. We found that the small geometric distortion of the blur area⁴ in the normalized camera space is tolerable and leads to practical focus estimations. Inconsistent intensity distributions would be far more critical. Thus, the normalized intensities f are used to estimate the focus values as explained in Section 5.

4.3 Sampling the Entire Surface

Each sample point in the camera view corresponds to a distinct area on the display surface. We estimate the focus of every projector pixel within the same surface area (i.e., the same sample point in the camera view).

To do this for the entire surface, multiple sample points are shifted within two-dimensional scan windows in the entire camera view.

The dimension of the scan windows is uniform, and depends on the resolution of the sample grid and the camera resolution. On the one hand, the grid's resolution should be coarse enough to avoid an interference of neighboring sample points (they have to be distinguishable in the camera image). On the other hand, it should be as dense as possible to speed up the scanning processes (since multiple sample points in the same frame can be processed in parallel). The radius of the sample points depends on the resolution of the projector and the camera, as well as on the distances of these devices to the projection surface and their field of view. We adjust these two parameters manually before the scanning process starts. They can also be determined automatically for a particular camera/projector configuration during an initialization step: The grid resolution can be successively decreased while the

3. And, the other intensities accordingly.

4. The actual sample points are geometry corrected.

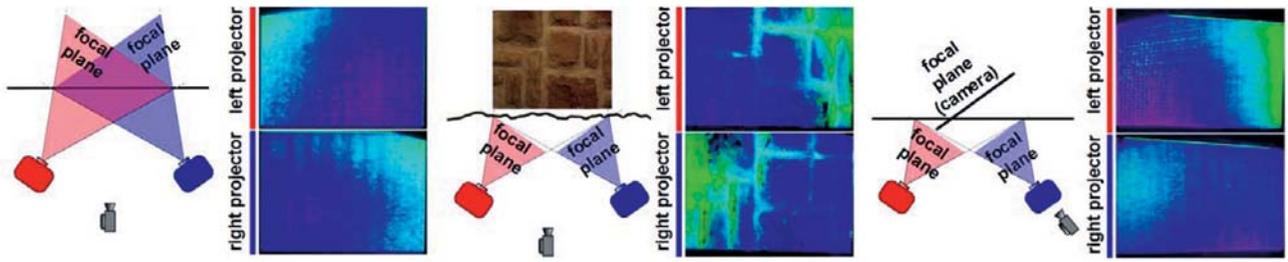


Fig. 5. Focus values measured for two projectors on different surfaces: white plane (left and right) and natural stone wall (center). The focus values have been color coded and have then been projected onto the surface (with correct geometry and radiometry). The focal plane of the camera approximates the average surface plane (left and right) or is parallel to one projector's focal plane (right). Color coding: blue to green = best to worst.

sample points' radii can be increased until the highest resolution and the smallest radii are found for which all sample points can still be detected and differentiated in the camera image.

A pixel-by-pixel shift of the grid mask results in exact focus estimations for every single camera pixel. The projector-individual focus values computed for a sample point that is larger than a camera pixel are mapped to its center pixel. Alternatively, the same focus values can be mapped to all pixels of the sample point. This allows shifts in the size of the sample points' radii and leads to shorter scanning times with lower precision. A third alternative is to map the focus values to the center pixel while performing larger shifts. The resulting voids can then be interpolated. This leads to even shorter scanning times, but also to a further reduction of the precision. For a planar surface, for example, it is sufficient to capture only one image with a coarse grid of sample points.

Note that all results presented in this article have been achieved with off-the-shelf hardware components, such as conventional video projectors and PAL-resolution camcorders.

Projecting the corrected sample points onto an extremely dark surface can cause radiances that are below the minimal camera response. In this case, the measurements are unreliable. To solve this problem, the grid of white sample points on a black background can be inverted before being corrected and projected. This results in a grid of black "sample holes" on a bright background (cf. Fig. 4). After recording, the camera image is inverted again to make the image compatible with the focus estimation functions (Section 5). The bright background leads to better focus estimations even with low camera response. The reason for this effect is illustrated in Fig. 4c: The projected pixels of the bright background are defocused at the surface. Their blur regions partially overlap which leads to an overall accumulation of radiances and consequently to an increase of intensities captured by the camera. These intensities are now above the minimal camera response and boundaries to darker areas are well detectable. Small defocused sample points, however, might not benefit sufficiently from an accumulation of radiances, and their intensities may fall below the minimal camera response—even if they are projected with the same intensity as the background pixels of the hole mask.

5 ESTIMATING RELATIVE FOCUS VALUES

Three general ways of deriving a focus value from a normalized intensity spread exist:

1. evaluating its intensity loss (which results from the distribution of constant light energy on a larger area),
2. measuring its frequency shift (since optical defocus has the same effect as the application of a digital low-pass filter), and
3. observing the enlargement of its area (which increases with increasing defocus).

The intensity loss can be computed by subtracting the maximum of the intensity spread from the corresponding value in the measured *FM* texture (see Section 3.1). Therefore, the intensity spread must be geometry corrected only. The radiometric compensation and postcorrection steps have to be disabled to lead to correct discrepancies. We found that focus operators that are based on intensity loss do not provide the demanded accuracy and were too noisy.

Frequency-based focus operators, such as Laplacian, discrete cosine transformation (DCT), or fast Fourier transformation (FFT), are often referred to in related literature. We applied and evaluated several such operators in the context of our problem. Most of them proven to deliver a relatively stable output and acceptable results in all situations. However, they are too expensive to compute when being applied to the search windows of every camera pixel's sample point.

Our experiments revealed that even better results can be received using the focus operator described in [23], which is originally used to measure the blur on CRT screens. Using the momentum preserving principle, this operator segments any discrete geometry of the intensity spread inside the corresponding search window into its fore and background. The proportion of foreground relative to the background is used as the focus value. This operator is fast and robust against camera noise. With an average computation time of 5.64s for a PAL-resolution image it is—in our experiments—by a factor of 3-11 faster than the evaluated frequency-based methods. Fig. 5 shows focus values measured with this operator on different surface types, and under different projector and camera settings. It demonstrates the feasibility of computing correct focus values independent of the projectors' or camera's position, or the surface properties.

Fig. 6 illustrates a single projector/camera configuration with a varying focal plane of the projector. The projection surface has a cylindrical shape. The measured focus values are visualized for each focal plane.

Note that the asymmetry along the vertical axis results from the vertical off-axis alignment of the projection frustum. Therefore, the image is always defocused more at the top than

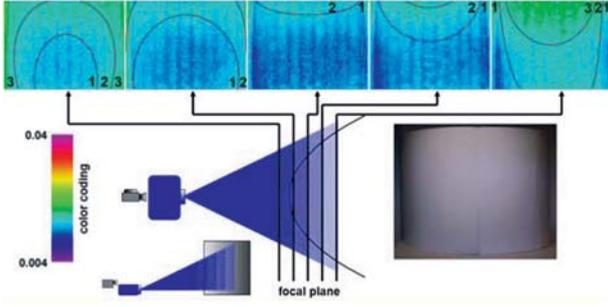


Fig. 6. Focus values measured for different focal plane positions on cylindrical surface. Color coding: blue to green = best (1) to worst (3).

at the bottom. The circular shape of the focus areas are due to slight radial focus variations caused by the projector’s lens systems and due to the cylindrical shape of the surface. It can be clearly seen that focused areas shift from the center to the outer sides as the focal plane is moved outward.

The focal plane of the camera causes additional blur and influences the absolute focus values. However, since it remains constant for all projectors the same amount of defocus is added in each measurement. Note that if the defocus of the camera is too high, small focus variations between the projectors might not be detectable with the limited resolution, intensity response and dynamic range of the camera. Consequently, the camera’s focus should be adjusted adequately to avoid extreme blur effects in the camera image. This allows computing relative focus values Φ that enable a quantitative comparison of corresponding projector pixels.

The focus values are first determined for each color channel separately and the results are averaged. The momentum preserving operator in [23] proven also to be stable among the RGB channels.

6 IMAGE COMPOSITION

The relative focus values $\Phi_{i,x,y}$ of each projector i that reaches a surface area which is visible in the camera pixel x, y are now known. Thus, we can compose an image from multiple projector contributions with minimal defocus in real-time. Two general techniques are imaginable (cf. Fig. 7): an exclusive composition or a weighted composition. Note that all projectors have to be photometrically calibrated to

produce an image with an overall consistent color and brightness. We apply well-known chrominance mapping and luminance matching techniques to achieve this. The description of these techniques is out of the scope of this article. A good overview can be found in [4].

6.1 Exclusive Composition

An exclusive image composition allows only one projector (the one with the largest focus value) to cover a surface area which appears at pixel x, y in camera space:

$$I_i = w_i(R - EM)/FM_i, w_i = \begin{cases} 1 & \Phi_{i,x,y} \geq \Phi_{j,x,y} \\ 0 & \text{else.} \end{cases} \quad (9)$$

The binary weights w_i can be determined offline from the focus values Φ , can be coded into a single texture map and can be passed to the projector-individual pixel shaders. Alternatively, a stencil mask can be computed for each projector. The mapping from camera space to projector space is given by the $C2P$ maps.

Neighboring pixels of different projectors might not be correctly aligned on the surface. This is due to their potentially unequal sizes and orientations, as well as due to imprecision of the geometric calibration. The resulting overlaps and gaps can lead to visible artifacts. To reduce these artifacts, the weight texture must be smoothed using a low-pass filter. This results in soft edges and in non-binary weight values. Since neighboring pixel-contributions from different projectors can now overlap partially, a different composition method has to be applied to ensure a correct radiometric compensation and blending. The weighted composition described below can be used for this as well.

6.2 Weighted Composition

Another disadvantage of the exclusive composition method is that the total light intensity that arrives at the surface cannot be larger than that produced by a single projector. This causes visual artifacts at surface patches with extremely low reflectance or bright environment light.

As explained in Section 3.2, the simultaneous contribution of multiple projectors can overcome this problem. A weighted image composition represents a tradeoff between intensity enhancement and focus refinement (cf. Fig. 8):

$$I_i = \frac{w_i(R - EM)}{\sum_j w_j FM_j}, w_{i,x,y} = \frac{\Phi_{i,x,y}}{\sum_j \Phi_{j,x,y}}. \quad (10)$$

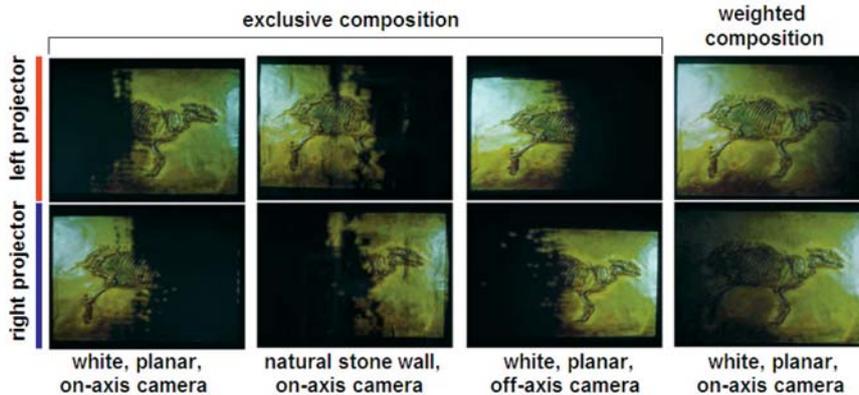


Fig. 7. Different image composition examples based on the configurations and estimated focus values shown in Fig. 5.

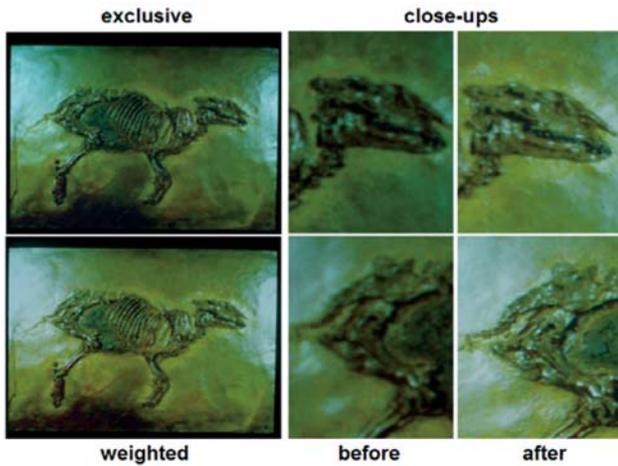


Fig. 8 Final composed image from contributions shown in Fig. 7: exclusive (top-left) and weighted (bottom-left). Close-ups: without (center row) and with (right row) multifocal projection.

The intensity contribution and the form factor component of each projector i that covers the same surface area x, y in camera space is weighted. The weights w_i are derived from the focus values and are normalized. They are not binary in this case. Projector contributions with high focus values are up-weighted and contributions of projectors with low focus values are down-weighted. Together, however, all contributions will always produce the correct result R when being radiated by the surface.

The weighted composition allows also to dynamically scale the focus values Φ_i up or down during runtime. This makes it possible to amplify or attenuate the contribution (and, consequently, the focus properties) of an individual projector i under retention of a correct radiometric compensation.

As described above, a softened exclusive weight texture avoids visible artifacts that result from pixel misregistrations on the surface. Only the weighted intensity (10) will ensure a correct radiometric compensation for this case. Instead of computing normalized weights directly from the focus values, the softened weights of the exclusive weight texture are normalized.

Similar as for the exclusive composition, a static alpha mask can be used alternatively for blending each projector's output instead of weighing the result inside the pixel shaders. Fig. 8 shows a final projection using the exclusive and the weighted composition methods based on the configurations illustrated in Fig. 5. Fig. 9 demonstrates a final projection via the exclusive composition based on the configuration shown in Fig. 6, and for a three-sided room corner. Fig. 10 illustrates another example of four projectors creating a focused image on a series of wallpapered surfaces within a depth-range of 1.5 meters.

In all three examples, a clear improvement in focus can be observed. Note that with simple photographs it is not possible to illustrate the overall improvement over the entire surface in a single image. The reason for this is that an adequately large aperture had to be applied to capture bright images which show the results. In this case, focus of our digital camera could be adjusted only for a relatively small area of the projection surface. It was then out of focus for other areas. The human

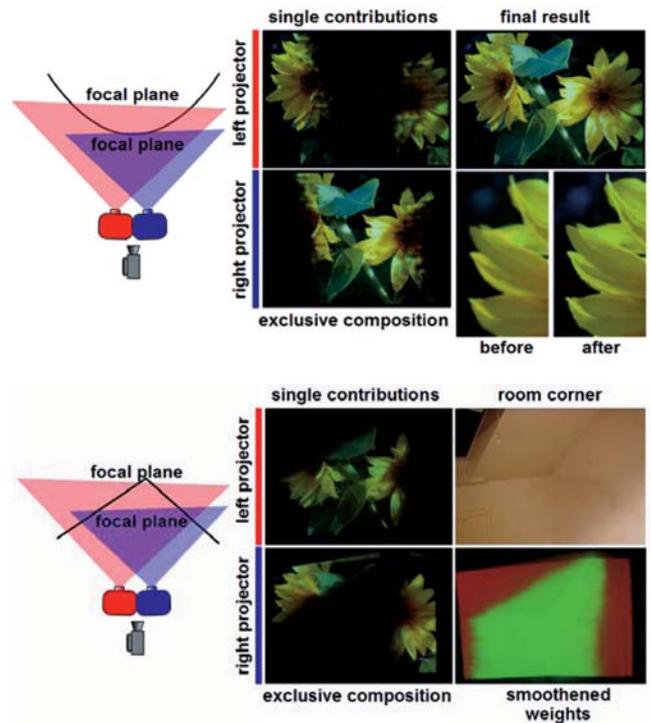


Fig. 9. Multifocal projection onto curved surface shown in Fig. 6 (top), and onto three-sided room corner (bottom). Both examples use two focal planes only.

visual system, however, adapts accommodation automatically when looking at different surface portions.

7 SUMMARY AND DISCUSSION

Many multiprojector techniques exist that aim at creating a high consistency of image *geometry*, *intensity*, and *color*. In this paper, we proposed a concept and a solution for considering and optimizing a fourth image property—its *focus*.

We described a method that aims at projecting images with minimal defocus onto geometrically complex, colored and textured surfaces. Multiple conventional projectors with overlapping images but differently adjusted focal planes are applied. Today's projectors already offer built-in autofocus mechanisms. Motorized adjustments of the lens system can be used to initially align the focal planes of each projector automatically. During an autonomous one-time calibration process, structured light projection together with camera feedback allows to measure the relative focus value of every projector pixel on an arbitrary diffuse surface. Thereby, the focus values are geometrically and radiometrically corrected. Thus, the measurements become independent from the properties of the projectors, the camera or the surface. Using the relative focus values, images with minimal defocus can be composed and displayed from individual pixel contributions of multiple projectors. Supported by hardware-accelerated pixel-shading, this can be performed in real-time.

Novel projection systems that support displaying images on everyday surfaces which are not optimized for projection represent a great potential for new applications in multimedia, information visualization, virtual reality, and augmented reality. Stereoscopic projection systems, for

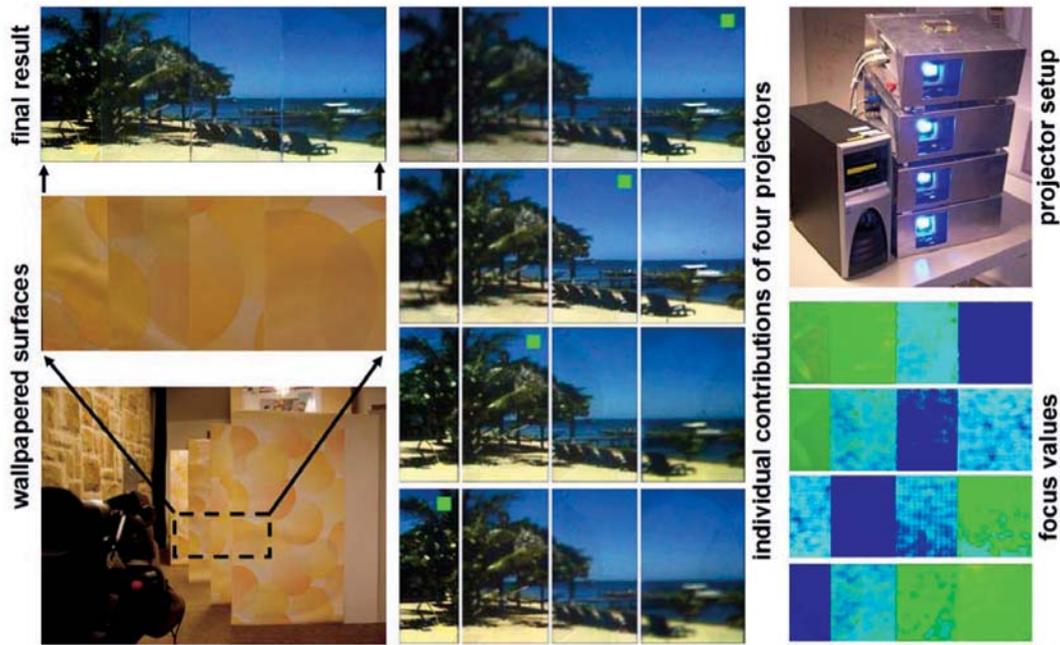


Fig. 10. Multifocal projection with four projectors onto consecutively aligned wallpapered surfaces (bottom left). The individual projector contributions (center) show their focused (highlighted with green square) and defocused areas on each surface. The final result (top-left) presents a focused image over the entire depth-range. All projections are geometrically corrected and radiometrically compensated. Focus values (bottom right) are color coded as before (blue to green = best to worst).

instance, that display images onto complex 3D surfaces with varying depth (e.g., [21], [15], [3]) can profit from a focus optimization. Otherwise, regional blur areas wash out image features which are needed for correct disparity-based depth perception. Stereo pairs become more difficult or even impossible to fuse in such situations.

Existing immersive projection technologies that employ cylindrical, spherical, or planar projection screens will not benefit much from a multifocal projection. For such parametric surfaces, optimal (or nearly optimal) projector placements and optical adjustments can be found (cf. Fig. 11, for example). But, for many display configurations, such as planetariums, the optimal projector placements cannot be realized because of structural or visibility constraints. A multifocal projection can still be applied for such displays to

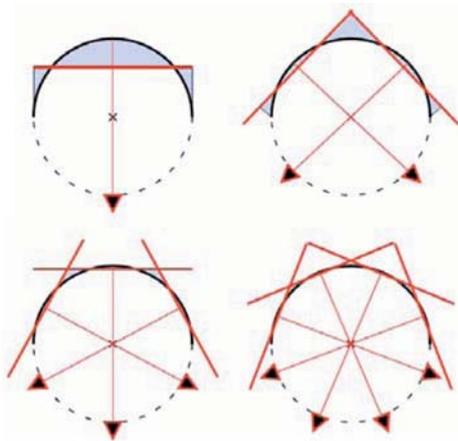


Fig. 11. Example for an optimal alignment of 1, 2, 3, and 4 projectors for spherical / cylindrical screen surfaces that minimizes defocus (illustrated in light blue).

seamlessly blend overlapping image portions via an automatic calibration. Today, simple cross-fading techniques are realized with linear ramps along the image planes that are proportional to the amount of image overlap. This is sufficient for planar screens, but not necessarily for curved screens. Deriving the blending function from defocus values, however, makes it proportional to the distance between focal planes and projection screen. This allows taking the screen surface into account, and to blend multiple projectors without manually registering their relative image geometries.

Such optimal configurations cannot be found for unconventional projection surfaces that are present at, for instance, airplane or car cabins, theme park installations, museums, or historical sites. Consequently, a multifocal projection together with unstructured projector alignments holds a larger potential in these cases.

The fact that multiple projectors are required to produce a single image without gaining in spatial resolution gives the impression of being uneconomic. However, we believe that for many applications an increased brightness and dynamic range, or an optimized focus can be equally important to a high resolution.

Obviously, the combination of multiple projection units allows improvements of most of these parameters. Thus, the amount of image overlap determines which of these properties is favored. A higher dynamic range, however, cannot be achieved by overlapping multiple images as long as projectors have a nonzero back-level.

The rapid development of digital projection technology indicates a continuous and fast improvement of resolution, brightness and dynamic range under falling costs. An increase of conventional projection systems' focal depth, however, is contrary to its enhancement in brightness and large throw ratio (= throw distance/projected image size). An extreme increase of focal depth through small apertures

has also to face optical limitations. The ongoing trend towards miniaturization and low-cost makes configurations such as the one illustrated in Fig. 1b well imaginable.

The high cost and the large proportions of emissive (e.g., laser or CRT) projectors will still reserve them for professional and special application domains, such as planetariums. However, the trend toward digital light-valve projectors will affect these areas sooner or later.

Future digital projector generations that are based on noncoherent light sources and on raster or scanning technologies can benefit from the advances of tiny liquid lenses to support a pixel-individual focus optimization. Yet, they need the capability to estimate per-pixel focus values automatically, rapidly, and independent of external conditions.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Armin Leitel from the Display Technologies Division of the Carl Zeiss Jena GmbH for his support and feedback. This project is funded by the Deutsche Forschungsgemeinschaft (DFG) under contract number PE 1183/1-1. To view the supplemental video please see <http://computer.org/tvcg/archives.htm>.

REFERENCES

- [1] W. Biehling, C. Deter, S. Dube, B. Hill, S. Helling, K. Isakovic, S. Klose, and K. Schiewe, "LaserCave—Some Building Blocks for Immersive Screens," *Proc. Int'l Status Conf. Virtual and Augmented Reality*, 2004.
- [2] O. Bimber, A. Emmerling, and T. Klemmer, "Embedded Entertainment with Smart Projectors," *Computer*, pp. 56-63, vol. 38, no. 1, 2005.
- [3] O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke, "Enabling View-Dependent Stereoscopic Projection in Real Environments," *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality*, pp. 14-23, 2005.
- [4] M. Brown, A. Majumder, and R. Yang, "Camera-Based Calibration Techniques for Seamless Multiprojector Displays," *IEEE Trans. Visualization and Computer Graphics*, vol. 11, no. 2, pp. 193-206, Mar.-Apr. 2005.
- [5] M.Ch. Chiang and T.E. Boulton, "Local Blur Estimation and Super-Resolution," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 821-826, 1997.
- [6] H. Eltoukhy and S. Kavusi, "A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction," *Proc. SPIE Electronic Imaging Conf.*, 2003.
- [7] K. Fujii, M.D. Grossberg, and S.K. Nayar, "A Projector-Camera System with Real-Time Photometric Adaptation for Dynamic Environments," *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 20-25, 2005.
- [8] P. Grossmann, "Depth from Focus," *Pattern Recognition Letters*, vol. 5, pp. 63-69, 1987.
- [9] M.D. Grossberg, H. Peri, S.K. Nayar, and P. Belhumeur, "Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 452-459, 2004.
- [10] P. Hill, N. Canagarajah, and D. Bull, "Image Fusion Using Complex Wavelets," *Proc. 13th British Machine Vision Conf.*, 2002.
- [11] C. Jaynes, S. Webb, and R.M. Steele, "Camera-Based Detection and Removal of Shadows from Interactive Multiprojector Displays," *IEEE Trans. Visualization and Computer Graphics*, vol. 10, no. 3, pp. 290-301, May-June 2004.
- [12] S.K. Kim, S.R. Park, and J.K. Paik, "Simultaneous Out-of-Focus Blur Estimation and Restoration for Digital Auto-Focusing System," *IEEE Trans. Consumer Electronics*, vol. 44, pp. 1071-1075, 1998.
- [13] S.K. Kim, S. Hwang, J. Shin, P.K. Paik, B. Abidi, and M.A. Abidi, "Object-Based Image Restoration for Multilayer Auto-Focusing," *Proc. SPIE-IS&T Electronic Imaging Conf.*, vol. 5308, pp. 893-901, 2004.
- [14] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic Aperture Confocal Imaging," *Proc. ACM SIGGRAPH '04*, pp. 825-834, 2004.
- [15] K.-L. Low, G. Welch, A. Lastra, and H. Fuchs, "Life-Sized Projector-Based Dioramas," *Proc. Symp. Virtual Reality Software and Technology*, pp. 93-101, 2001.
- [16] A. Majumder and G. Welch, "Computer Graphics Optique: Optical Superposition of Projected Computer Graphics," *Proc. Eurographics Workshop Virtual Environment/Immersive Projection Technology*, pp. 209-218, 2001.
- [17] M. McGuire, W. Matusik, H. Pfister, J.F. Hughes, and F. Durand, "Defocus Video Matting," *Proc. ACM SIGGRAPH '05*, pp. 567-676, 2005.
- [18] S.K. Nayar, M. Watanabe, and M. Noguchi, "Real-Time Focus Range Sensor," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1186-1198, 1996.
- [19] S.K. Nayar, H. Peri, M.D. Grossberg, and P.N. Belhumeur, "A Projection System with Radiometric Compensation for Screen Imperfections," *Proc. Int'l Workshop Projector-Camera Systems*, 2003.
- [20] M. Noguchi and S.K. Nayar, "Microscopic Shape from Focus Using Active Illumination," *Proc. Int'l Conf. Pattern Recognition*, pp. 147-152, 1994.
- [21] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays," *Proc. ACM SIGGRAPH '98*, pp. 179-188, 1998.
- [22] R. Raskar, J. van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines, "iLamps: Geometrically Aware and Self-Configuring Projectors," *Proc. ACM SIGGRAPH*, pp. 809-818, 2003.
- [23] D.M. Tsai and C.C. Chou, "A Fast Focus Measure for Video Display Inspection," *Machine Vision and Applications*, vol. 14, no. 3, pp. 192-196, 2003.
- [24] D. Wang, I. Sato, T. Okabe, and Y. Sato, "Radiometric Compensation in a Projector-Camera System Based on the Properties of Human Vision System," *Proc. IEEE Int'l Workshop Projector-Camera Systems*, 2005.



Oliver Bimber received the PhD degree in engineering from Darmstadt University of Technology, Germany. He is a junior professor of augmented reality at Bauhaus University, Weimar, Germany. His research interests include display technologies, computer graphics, and computer vision. He is a member of the IEEE and the IEEE Computer Society



Andreas Emmerling graduated with a major in media systems from Bauhaus University Weimar in June 2005. His research interests include projection technology, computer graphics, and computer vision.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.

Projects in VR

Editors: Larry Rosenblum and
Simon Julier

Augmenting Holograms

Oliver Bimber
Bauhaus-
University
Weimar

Among all the imaging techniques invented throughout the past several decades, computer graphics represent one of today's most successful methods. Many areas in science, entertainment, education, and engineering would be unimaginable without the aid of 2D or 3D computer graphics. Interactivity provides an important reason for this success story, a property that competing technologies—such as holography—still cannot provide efficiently.

In parallel to the development of computer graphics, and despite their noninteractivity, advances in optical and digital holography have made possible the creation of new fields including interferometry, copy protection, data storage, holographic optical elements, and display holograms. Display holography in particular has conquered several application domains. Museum exhibits often use optical holograms because they can present 3D objects with almost no loss in visual quality. In contrast to most stereoscopic or autostereoscopic graphics displays, holographic images can provide all depth cues.

Today, computer graphics and raster displays offer megapixel resolution and interactive rendering of megabytes of data. Optical holograms, however, provide a terapixel resolution and can present information content in the range of terabytes in real time. Both dimensions will be out of reach for computer graphics and conventional displays for the next several years.

Researchers in several domains are using holography and computer graphics as tools to solve individual research, engineering, and presentation problems. Up until now, however, these tools have been applied separately.

The goal of the HoloGraphics project at Bauhaus-University Weimar was to investigate possibilities of combining the advantages of conventional holograms—such as high visual quality and realism, support for all depth queues and for multiple observers at no computational cost, and space efficiency—with the advantages of today's computer graphics capabilities—such as interactivity, real-time rendering, simulation and animation, and stereoscopic and autostereoscopic presentation. This article presents some of our project's findings.

Reconstructing holograms with digital light

The two basic hologram types—transmission and reflection—are both reconstructed by illumination with a reference light wave. These two types have generated

a number of variations. Although we can only reconstruct some holograms with coherent laser light, we can view others under white light.

Conventional video projectors represent point sources well suited for viewing white-light reflection or transmission holograms. Projectors offer an advantage over analog lightbulbs because we can digitize the reference wave used to reconstruct the hologram. This makes it possible to control the amplitude and wavelength of each discrete portion of the wavefront over time.

When integrating graphical elements into an optical hologram, reconstructing the holographic image only partially using digital light is key.¹ Our system uses the holographic film itself as the optical combiner, since it remains transparent if not illuminated from the correct angle. We place raster displays behind the film, using projected images to replay the hologram in a controlled way.

Using stereoscopic or autostereoscopic displays to render 3D graphics registered to a hologram allows both holographic and graphical content to appear as three-dimensional within the same space.

Using projected light, we can reconstruct a hologram only partially, leaving gaps where graphical elements can be inserted. A user can then view interactive computer graphics through these gaps simultaneously with the holographic content. In this way, rendering and illumination can become view-dependent if they are synchronized.

The replayed hologram's intensity is proportional to the intensity of the reconstruction light. In addition to using an incomplete illumination for reconstructing a fraction of a hologram, intensity variations of the projected light permit local modification of its brightness.

Having depth information of both holographic and graphical content allows for the computation of an illumination image, which reconstructs a hologram in such a way that graphical components can be integrated consistently. This includes correct occlusion and shading effects.

Figure 1 illustrates a desktop display prototype. It uses an autostereoscopic lenticular-lens sheet display with integrated head finder for wireless user tracking and a force feedback device for 6-degrees-of-freedom interaction. A digital light projector illuminates the hologram from the correct angle. A single PC with a dual-output graphics card renders the graphical content onto the screen and the illumination for the holographic film onto the video projector. The holographic film is mounted in front of the screen.

Figure 2a visualizes depth information reconstructed from a 30 × 40-centimeter rainbow hologram of a dinosaur skull created using a flatbed scanner.² The depth information is registered to the skull's holographic image to allow the computation of consistent occlusion and shading effects.

Figures 2b and 2c illustrate the holographic image of the skull, combined with graphical representations of soft tissue and bones. We can create synthetic shading and shadowing effects, as well as correct occlusions, in real time from different positions of an interactive virtual light source and for an arbitrary point of view.

Augmenting large-scale holograms

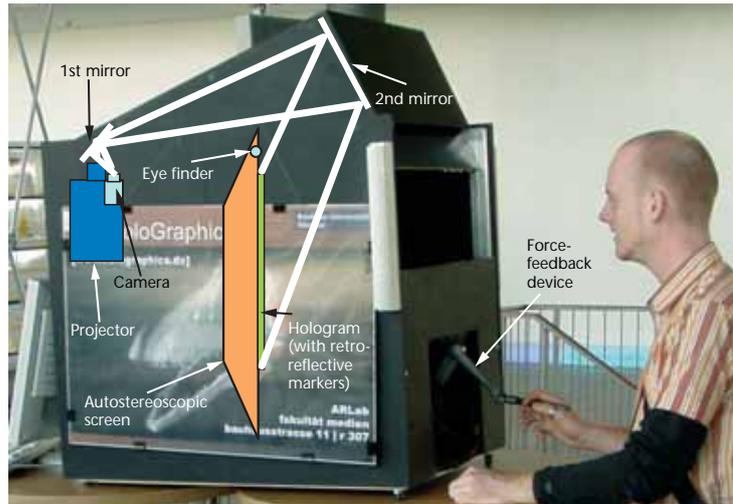
Large-scale optical holograms require large-scale display technology for augmentation with interactive graphical elements. Conventional LCD or CRT desktop displays are not sufficient for this task.

We can use projectors instead for both illuminating and augmenting large-scale holograms. Projecting stereoscopic images directly onto the holographic grating, however, has one main disadvantage: of the incoming light transmitted, only a small fraction is diffused. This inefficient process leads to relatively dim images. Although more efficient projection screens are available, these screens must not be opaque and have to transmit and diffuse light for replaying holograms and for displaying graphics.

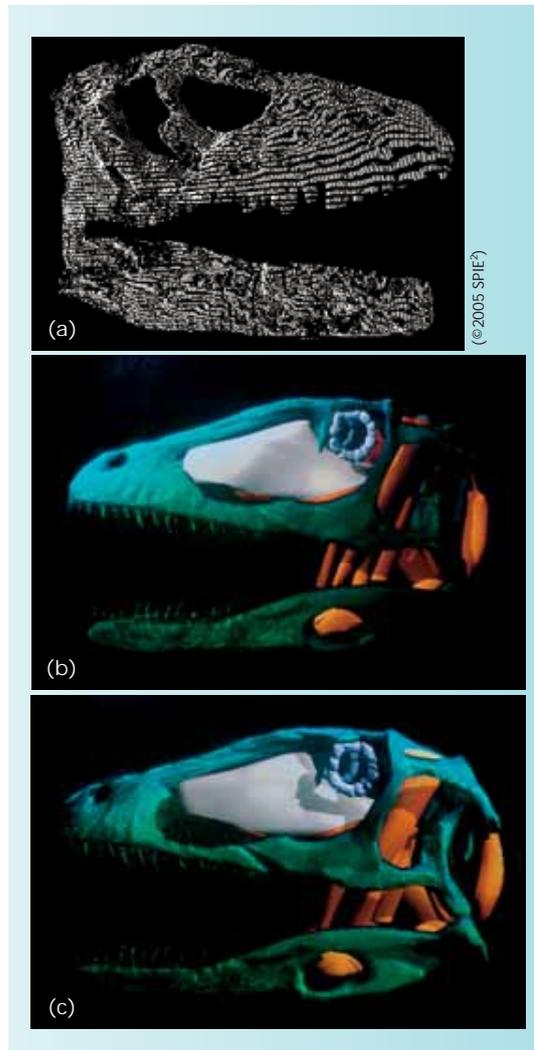
Examples of passive projection options include semi-transparent holographic projection screens that diffuse the light in a narrow angle to achieve an enhanced brightness for restricted viewing angles. Transparent film screens that diffuse the light in a wide angle could also support more flexible viewing and a higher degree of transparency. Other possibilities include actively shuttered projection screens with a phase-dispersed liquid crystal layer (PDLC) that we can electronically switch to a diffuse or transparent state.

Figure 3a (next page) illustrates our configuration that applies a large PDLC screen to integrate stereoscopic graphics into a 170 × 103-cm large rainbow hologram of a *Tyrannosaurus rex* skull (see Figure 3b). We scanned the hologram using a two-lens stereo camera system (see Figure 3c) and reconstructed it from multiple merged point clouds (see Figure 3d). We mounted the PDLC screen behind the holographic film. While an illumination projector reconstructs the hologram in the transparent state of the PDLC, a display projector generates the stereoscopic images of the augmented graphical content in the diffuse state. The left and right stereo images of the graphical content are displayed with twice the frequency (in this case, 100 Hz) of the PDLC screen's shutter frequency (50 Hz). Because the display involves active stereo projection, the head-tracked observer wears shutter glasses.

Figures 4a and 4b show the *T. rex* skull augmented with an animated finite-element method simulation of cranial mechanics and feeding. Figures 4c-e illustrate the optical integration process: without controlled reconstruction of the holographic content, the integrated graphical content appears semitransparent (see Figure 4c). Not reconstructing the holographic content

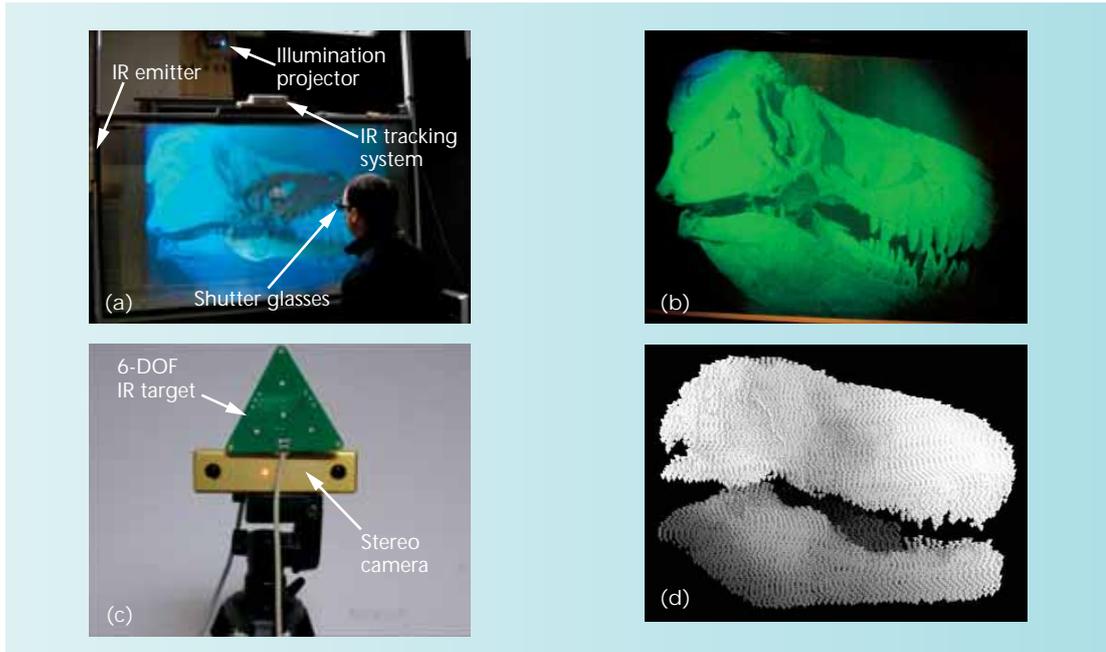


1 Autostereoscopic desktop display prototype for augmenting a hologram of a dinosaur skull with rendered soft tissue.

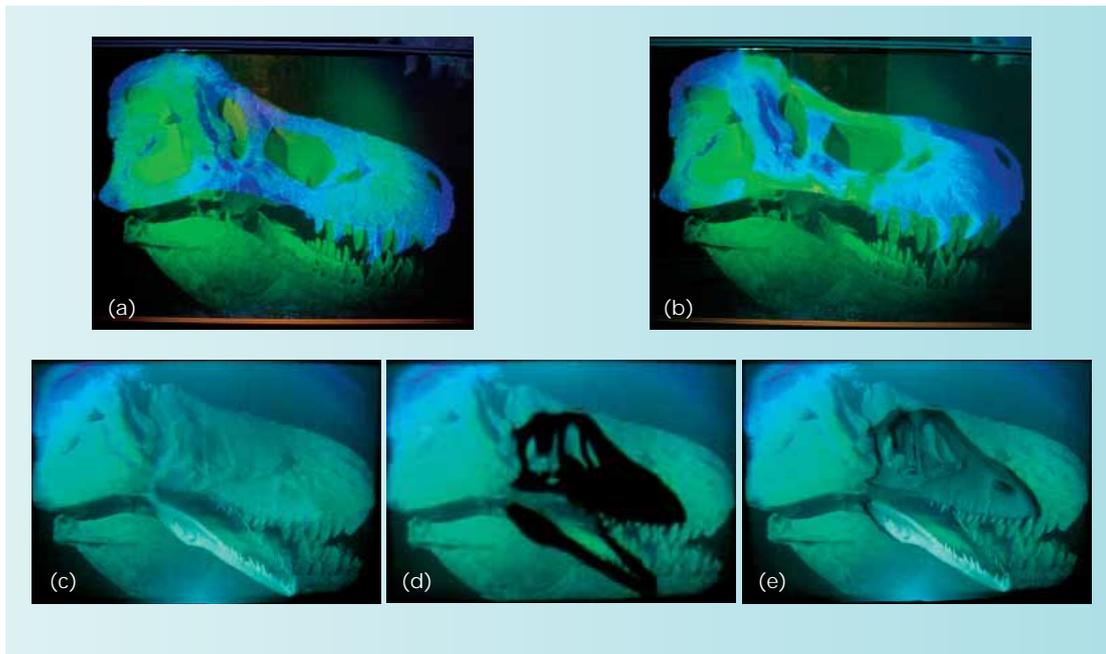


2 We need the (a) depth of the hologram to compute (b, c) consistent occlusion and shading effects.

3 (a) A projection onto a phase-dispersed liquid crystal screen augments a (b) large hologram of a Tyrannosaurus rex skull. (c) A two-lens range scanner captures multiple merged point clouds to reconstruct (d) the holographic depth.



4 (a-e) Registered to the holographic content, the depth information ensures a consistent augmentation.



in areas where graphical elements occlude holographic elements (see Figure 4d) leads to consistent occlusion effects (see Figure 4e).

The holographic content also partially occludes the graphical content. As explained earlier, this requires registered depth information of the hologram.

Conventional flatbed scanners cannot reconstruct depth information from large holograms. Range scanning might seem like a possible efficient alternative, but for small holograms, due to the scanners' low precision and the holograms' limited field of view, this approach is impractical. Thus, for depth reconstruction we used flatbed scanners for small holograms and range scanners for large holograms.

Since the cameras in the two-lens range scanning system we used have a limited field of view and the holographic content has to be reconstructed from multiple perspectives, we use multiple scans to capture different horizontal angles. (A rainbow hologram does not contain a vertical parallax.) Each scan results in a 3D point cloud of approximately 3,500 to 7,000 points on different fractions of the skull's surface.

To form the whole surface, we must merge the different point clouds into a common coordinate system. We achieve this by determining the range sensor's position and orientation relative to the holographic plate, using an infrared tracking system. By first aligning multiple fractional point clouds roughly with the transform-

mation provided by the tracking system and then matching them using numerical closest-point iterations, we could register them into a common coordinate system.

Due to the limited resolution of the range sensor, small gaps appear between the actual surface points. To obtain continuous depth information, the points could be triangulated. This leads to a connected triangle mesh that allows us to estimate the missing depth information via interpolation.

Triangulation of such an unstructured point cloud, however, poses problems. Because of the complex topology of the surface, automatic triangulation methods will most likely create wrong connectivities between points. This leads to a significant amount of manual postprocessing. Instead of triangulating the points into a mesh of triangle primitives, we leave the points unconnected. They are rendered as point primitives with appropriate radii to fill the gaps of missing surface information (see Figure 3d).

Dynamic multiresolution splatting techniques provide fast and adaptable frame rates as well as a continuous representation of a discretized surface.³ Normally the splats are not visualized but instead are only registered to the holographic content for computing consistent occlusion and shading effects.

Augmentation techniques for holograms

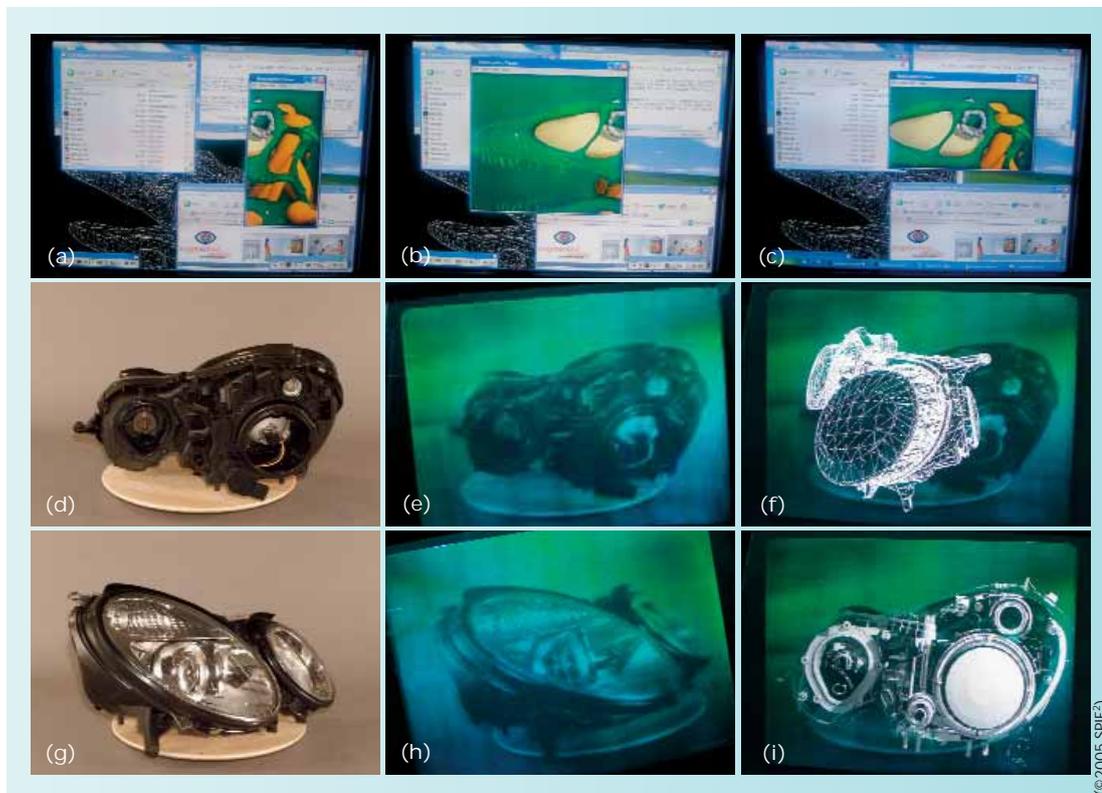
The ability to digitally control the reconstruction of a hologram allows a user to integrate it seamlessly into common desktop window systems. If the holographic film mounted in front of a screen is not illuminated, it remains transparent. In our example, the entire screen content is

visible, and interaction with other software applications on the desktop is possible. The holographic content (visible or not) is always located at a fixed spatial position within the screen or desktop reference frame.

An application that renders graphical content does not necessarily need to be displayed in a full screen mode but instead can run in a windows mode, covering an arbitrary area on the desktop behind the film. If we know the position and dimensions of the application window, we can synchronize the digital illumination to bind the projected light to the portion of the film located directly on top of the underlying window. This simple but effective technique allows a seamless integration of holograms into common desktop environments. It allows a user to temporarily minimize such a holographic window (see Figures 5a–c) or to align it over the main focus while working with other applications.

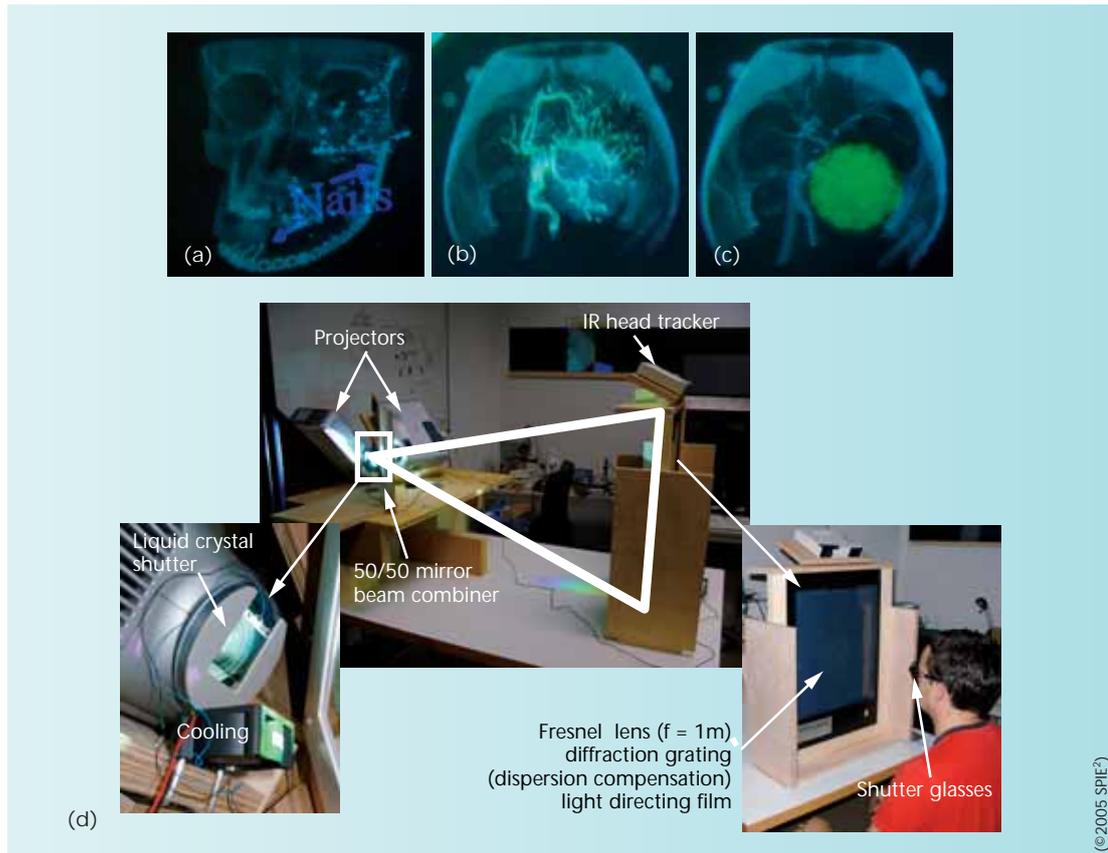
Holograms can be multiplexed: we can optically store multiple recordings of different information on the same film through several exposure steps. We can divide the entire viewing space—in theory about 180 degrees in horizontal and vertical directions, but in practice less, due to optical limitations—into different viewing zones that are recorded with individualized content, such as independent perspectives.

Figures 5d–i illustrate a multiplexed hologram of a car headlight augmented with CAD data. An observer can see three different partial views (front, rear, and side) by moving within the total viewing range of 110 degrees. Since the observer’s head motion is tracked and the recorded viewing angles are known, the system can update the graphical content perspective to match the corresponding perspective recorded in the hologram. This supports



5 (a–c) We can restrict a hologram augmentation to the dimensions of a holographic window on common desktop environments. (d–i) Multiplexing holograms allows an increase in the viewing range.

6 We can augment (a–c) volumetric multiplexed holograms by (d) projecting illumination and graphical content in one image.



a noncontinuous surround view of recorded and augmented scenes.

Volumetric multiplexed holograms⁴ are digital multiple exposure transmission holograms that contain static computerized tomography (CT), magnetic resonance imaging (MRI), or other volumetric data sets. These hologram systems encode slices from a medical scan on a holographic film. The slices can be reconstructed simultaneously and appear as semitransparent in space, producing 3D volumetric images. In contrast to many other stereoscopic or autostereoscopic display techniques used for viewing volumetric data, volumetric multiplexed holograms share all properties of other hologram types and support all depth queues—including correct focus.

We can augment the static volumetric multiplexed holograms with interactive stereoscopic graphics (see Figures 6a–c). In the display configuration shown in Figure 6d, one stereoscopic projector (or two optically aligned monoscopic projectors) displays the illumination and the colored stereo pairs simultaneously.

New applications

Merging optical holograms with interactive computer graphics could lead to new tools for science, industry, and education.

Archaeologists, for example, already use holograms to archive and investigate ancient artifacts. Scientists can use hologram copies to perform their research without having access to the original artifacts or settling for inaccurate replicas. They can combine these

holograms with interactive computer graphics to integrate real-time simulation data or perform experiments that require direct user interaction, such as packing reconstructed soft tissue into a fossilized dinosaur skull hologram. In addition, specialized interaction devices can simulate haptic feedback of holographic and graphical content while scientists are performing these interactive tasks.

An entire collection of artifacts will fit into a single album of holographic recordings, and researchers can use a light box display—such as those that present x-rays—for visualization and interaction. The same applies to the biomedical domain, which already uses digital volumetric holograms produced from CT or MRI data of inner organs.

In the automotive industry, complex computer models of cars and components often lack realism or don't support interactivity. Instead of attempting to achieve high visual quality and interactive frame rates for the entire model, designers could decompose the model into sets of interactive and static elements. The system could record physical counterparts of static elements in a hologram with maximum realism and release computational resources to render the interactive elements with a higher quality and increased frame rate. Multiplexing the holographic content also lets users observe and interact with the entire model from multiple perspectives.

Interferograms used for nondestructive measurement and testing are yet another example of industrial applications. Users can combine high-resolution analog

interferograms that indicate object motion, vibration, or deformation with digital simulation data.

Augmenting display holograms in museums with animated multimedia content lets exhibitors communicate information about the artifact with more excitement and effectiveness than text labels offer. Such displays can also respond to user interaction. Because wall-mounted variations require little space, museums can display a larger number of artifacts.

Optical holograms can store a massive amount of information on a thin holographic film. This technology can record and reconstruct a 3D scene with almost no loss in quality. Moore's law must be applied many times for us to reach this quality of graphic or electroholographic rendering techniques and displays at interactive frame rates. A combination of interactive computer graphics and high-quality holograms represents an alternative that we can realize today with off-the-shelf consumer hardware. ■

Acknowledgments

Special thanks to the HoloGraphics team at Bauhaus-University Weimar, whose work is presented in this article. Ohio University, Natural History Museum London, and DaimlerChrysler provided the data visualized in Figures 1, 2, and 3. The Deutsche Forschungsgemein-

schaft (DFG) supported this project under contract number BI 835/1-1. Further information about the project is available at <http://www.HoloGraphics.com>.

References

1. O. Bimber, "Combining Optical Holograms with Interactive Computer Graphics," *Computer*, Jan. 2004, pp. 85-91.
2. O. Bimber et al., "Interacting with Augmented Holograms," *Proc. SPIE Conf. Practical Holography XIX: Materials and Applications*, Springer, 2005.
3. S. Rusinkiewicz and M. Levoy, "QSplat: A Multiresolution Point Rendering System for Large Meshes," *Proc. ACM Siggraph*, ACM Press, 2000, pp. 343-352.
4. S.J. Hart and M.N. Dalton, "Display Holography for Medical Tomography," *Proc. SPIE, Practical Holography IV*, vol. 1212, Springer, 1990, pp. 116-135.

Readers may contact Oliver Bimber at obimber@computer.org.

Readers may contact the department editors at lrosenbl@nsf.gov and S.Julier@cs.ucl.ac.uk.

Merging Graphics and Holograms

Oliver Bimber
Bauhaus-University Weimar
Bauhausstr. 11, 99423 Weimar, Germany
Ph.:+49-(0)3643-583724 , Fx.:+49-(0)3643-583709
Email: bimber@uni-weimar.de

ABSTRACT

This article outlines how display holograms can be combined with interactive computer graphics. Digitally projected light is used for replaying the holographic content synchronized to the rendering of autostereoscopic or stereoscopic graphics. Modifying the local intensity of the projected light beam allows creating consistent occlusion and shading effects between both – graphics and hologram. This, however, requires depth information of the holographic recording. While flatbed scanners are suitable for estimating surface depth of small to medium size white-light holograms, range scanning is preferred for large-scale holograms. While the integrated graphical elements allow interactivity that is not supported by analog display holograms, the holograms can provide a visual quality that is not possible with today's three-dimensional displays.

1 INTRODUCTION

Many areas in science, entertainment, education, and engineering would be unimaginable without the aid of 2D or 3D computer graphics. The reason for the success story of computer graphics might be its interactivity, which is an important property that is still not provided efficiently by competing technologies – such as holography.

While display holography is limited to presenting a non-interactive content, electroholography or computer generated holograms (CGH) facilitate the computer-based generation and presentation of holograms at interactive rates [2,3,20,21]. Holographic fringes can be computed by either rendering multiple perspective images, then combining them into a stereogram [4], or by simulating the optical interference and calculating the interference pattern [5]. Special display systems dynamically generate the output wavefront from the computed fringe data. Since creating an electrohologram requires processing, transmitting, and storing a massive amount of data, today's computer technology still sets the limits for this technology. To overcome some of these performance and storage issues, advanced reduction and compression methods have been developed that create truly interactive electroholograms. Unfortunately, most of these holograms are relatively small, low resolution, and cover only a small color spectrum. However, recent advances in consumer graphics hardware may reveal potential acceleration possibilities that can overcome these limitations [6].

Especially display holography has conquered several public application domains, such as museums, theme parks and trade shows. Displaying artifacts virtually removes the need to build physical replicas or showcase originals. This can save display space. But the true reason for displaying holograms might mainly lie in the fascination of viewing three-dimensional images. In addition, holograms can be used to make engineering, medical, dental, archaeological, and other recordings—for teaching, training, experimentation and documentation. Archaeologists, for example, use optical holograms to archive and investigate ancient artifacts [7,8].

In contrast to most stereoscopic or autostereoscopic graphics displays, holographic images can provide all depth cues—perspective, binocular disparity, motion parallax, convergence, and accommodation—and theoretically can be viewed simultaneously from an unlimited

number of positions. Today, computer graphics and raster displays offer a megapixel resolution and the interactive rendering of megabytes of data. Optical holograms, however, provide a terapixel resolution and are able to present an information content in the range of terabytes in real-time. Both are dimensions that will not be reached by computer graphics and conventional displays within the next years.

Obviously, one has to make a decision between interactivity and quality when choosing a display technology for a particular application. While some applications require high visual realism and real-time presentation (that cannot be provided by computer graphics), others depend on interactivity (which is not possible with analog display holograms).

The intention of the *HoloGraphics* project¹ which is outlined in this article is to combine both technologies. Several possibilities have been investigated that allow merging computer generated graphics and white-light holograms [1]. The goal is to combine the advantages of conventional holograms (i.e. extremely high visual quality and realism, support for all depth queues and for multiple observers at no computational cost, space efficiency, etc.) with the advantages of today's computer graphics capabilities (i.e. interactivity, real-time rendering, simulation and animation, stereoscopic and autostereoscopic presentation, etc.).

2 REPLAYING OPTICAL HOLOGRAMS WITH DIGITALLY PROJECTED LIGHT

Conventional video projectors represent point sources that are well suited for viewing most white-light reflection or transmission holograms. The main advantage for using projectors instead of analog light bulbs -like halogen spots- is that the light beam used to replay a hologram can be dynamically digitized. Thus it is possible to control the intensity and color of each discrete portion of the reference beam over time.

Early experiments with video projectors for reconstructing optical holograms have been made in the art and engineering domains. In some art installations, optical holograms have been linked with time-based media, such as slides, film-loops or color patterns that are projected onto them to achieve artistic effects [9,10]. Others have redirected projected light with multiple mirrors to simulate different light sources. The goal was to achieve dynamic fluctuation effects with optical holograms [11,12].

Reconstructing a hologram only partially using a projected light beam is the key concept of integrating graphical elements into an optical hologram [1]. The holographic film itself can be used as optical combiner, since it remains transparent if not illuminated from the correct angle (cf. figure 1).

Raster displays can be placed behind the film while projected light can be used for replaying a hologram in a controlled way (cf. figure 2). If stereoscopic or autostereoscopic displays are used to render 3D graphics registered to a hologram, then both holographic and graphical content can appear three-dimensional within the same space. Continuous head-tracking is required for displaying the graphical content correctly to a single observer with stereoscopic or two-view autostereoscopic displays. Some stereoscopic display methods support multiple head-tracked observers simultaneously. Multi-view autostereoscopic displays, however, support many observers at the same time without the need for head-tracking.

A hybrid display approach has already been described earlier that combined a transmission hologram with a liquid crystal display to realize a new user interface for business machines, such as photocopiers [13]. An analog point light source was used to illuminate a transmission hologram which was mounted behind an LCD panel. In this case it was not possible to control the reconstruction of the holographic image at discrete areas.

Using projected light it is possible to reconstruct a hologram only partially, leaving gaps where graphical elements can be inserted. Interactive computer graphics can then be viewed

¹ www.HoloGraphics.de

through these gaps simultaneously with the holographic content. Thereby, rendering and illumination are view-dependent and have to be synchronized.

The replayed hologram's intensity is proportional to the intensity of the projected reference beam. In addition to using an incomplete illumination for replaying a fraction of the hologram, intensity variations of the projected light permit local modification of its brightness (i.e., the local brightness of the replayed image beam).

Having depth information of both –holographic and graphical content– allows computing a binary or shaded illumination image which replays the hologram in such a way that graphical components can be integrated consistently. This includes correct occlusion and shading effects (cf. figure 3).

Synthetic shading and shadowing effects between graphical and holographic content from different positions of an interactive virtual light source, as well as correct occlusions are created in real-time for an arbitrary point of view. While the hologram remains static, the three-dimensional computer graphics is fully interactive.

Large scale holograms require large scale display technology for merging them with interactive graphical elements. Conventional displays, such as desktop-size CRTs or LCDs are not sufficient in this case.

Projectors can be used instead for both – illuminating and augmenting large holograms. Projecting stereoscopic images directly onto the holographic grating, however, has one main disadvantage: Most of the incoming light is transmitted and only a small fraction is being diffused. This is inefficient and leads to relatively dim graphical images – but more efficient projection screens are available. These screens, however, must not be opaque, but have to transmit and diffuse light for replaying the hologram and for displaying the graphics simultaneously or sequentially with a high speed.

Examples for passive screens are semi-transparent holographic projection screens that diffuse the light in a narrow angle to achieve an enhanced brightness for a restricted viewing range, or transparent film screens that diffuse the light in a wide angle to support a more flexible viewing and a higher degree of transparency. Other possibilities are actively shuttered projection screens with a phase dispersed liquid crystal (PDLC) layer that can be switched to a diffuse and to a transparent state electronically.

Figure 4 illustrates a configuration that applies a large PDLC screen to integrate stereoscopic graphics into a 170cm x 103cm large rainbow hologram of a T.Rex skull (cf. figure 5). The PDLC screen is mounted behind the holographic film. While an illumination projector reconstructs the hologram from the required angle in the transparent state of the PDLC, a display projector generates the stereoscopic images of the augmented graphical content in the diffuse state. The left and right stereo images of the graphical content are displayed with twice the frequency (i.e., 100Hz in this case) of the screen's shutter frequency (i.e., 50Hz in this case). Due to the active stereo projection, the head-tracked observer wears shutter glasses.

Figure 5 shows the integration of graphical content into a large rainbow hologram with and without synchronized PDLC screen, as well as with and without controlled illumination. Note that due to a small gap between the PDLC screen and the holographic plate and due to the fact that the light is being scattered by the screen in all directions in its diffuse mode, the holographic content appears achromatic and slightly blurred (cf. figure 5-bottom row). This can be reduced by minimizing the distance between holographic film and diffuse screen.

3 RECONSTRUCTING DEPTH FROM HOLOGRAMS

Depth information of the recorded holographic content is essential for computing the correct illumination images that enable a consistent occlusion and shading with the graphical content. Once they are known, the depth values have to be registered exactly to holographic image and

to film plane. This is required for accomplishing accurate shading computations within the illumination images which are also registered to the holographic image and to the film plane. Several approaches for reconstructing depth information from optical holograms have been developed. Ultra-fast holographic cameras, for instance, have been modified to allow capturing 3D objects, such as faces [14] or bodies [15]. A fast pulsed laser with short exposure time is used in these cases for holographic recording that is free of motion artefacts. The depth information is then reconstructed by replaying the hologram with a laser. Topometric information is retrieved by digitizing the real holographic image that is projected onto a diffuse plate. Moving the plate in the depth direction (away from the holographic plate) results in several 2D slices through the holographic image. These slices are finally combined to form the corresponding 3D surface. Other approaches are based on the holographic light-in-flight technique [16]. They measure the shape of a recorded surface by determining the time for light to travel from different points of the object.

While the techniques outlined above reconstruct depth from laser-light holograms, the following methods summarize how this can be achieved for white-light holograms.

The depth of small to medium size white-light holograms can be reconstructed with conventional flatbed scanners [22]. Multiple images of the hologram are scanned by placing the holographic film on top of the scanner window, leaving the lid open and illuminating it under different illumination angles for each scan (cf. figure 6). The geometric image distortion that is caused by the different illumination angles has been described by Champagne [18]. Based on this model, a method has been derived by Bazargan [19] to estimate the image position of recorded objects for cases in which the recording reference beam does not match the reply reference beam². A modified version of this approach can be applied for computing the depth of each object point on the hologram's surface for which corresponding image points in at least two scan images are known [22] (cf. figure 6).

Reconstructing depth from large holograms is impossible with conventional flatbed scanners. For these hologram types, other techniques become more efficient. Range scanning is one possibility. Passive range sensors, for instance, rely on the input from two or more calibrated perspective cameras, and on correlating the images geometrically. In this case conventional two-view or multi-view geometry techniques can be used for reconstructing the depth, once the point correspondences in the perspective images have been determined.

Applying range scanning to small and medium size holograms, however, is impractical due to the sensors' low precision and the holograms' limited field of view: Scanning the content close to the hologram restricts the scanning angle of the sensor. Scanning the content from an adequate distance which permits a larger angular scanning range does not allow dissolving small features that are required for image correlation.

Figure 7 illustrates the depth information reconstructed from the large T.Rex rainbow hologram. A two-lens stereo camera system was used for scanning point clouds from multiple horizontal angles. The resulting point clouds that describe individual surface fractions are registered into a common coordinate system using numerical closest-point iterations [24]. To obtain a continuous depth map, the points could be triangulated. This leads to a connected triangle mesh that allows estimating the missing depth information via interpolation. Triangulation of such an unstructured point cloud, however, is difficult. Due to the complex topology of the point cloud, automatic triangulation methods will most likely create wrong connectivities between points. This leads to a significant amount of manual post processing. Instead of triangulating the points into a mesh of triangle primitives, the points remain unconnected. They are rendered as point primitives with appropriate radii to fill the gaps of missing surface information. Such a point-based rendering concept is referred to as dynamic multi-resolution splatting [23] and supports real-time computations.

² A known perspective projection is assumed.

4 SUMMARY

Dedicated scientists such as Leith, Upatnieks, Benton, and many more have taken over Gabor's legacy and helped to make out of holography what it is today: more than just fascinating three-dimensional images. Modern holography has created new fields including interferometry, copy protection, data storage, or holographic optical elements. So far and with a few exceptions, traditional display holography has found its main application in museums. This article has outlined how analog display holograms and interactive computer graphics can be combined. Using controllable digital light projection, consistent occlusion, shading and shadow effects between both components can be created. This has been achieved not only for rainbow holograms as illustrated in this article, but also for a variety of other hologram types, such as reflection holograms, full color holograms (such as [25]), digital multiplex stereograms, and volumetric multiplexed holograms [26]. The depth information that is essential for carrying out the corresponding computations can be reconstructed from small to large optical holograms. While holograms can provide a visual quality that is not possible with today's three-dimensional displays, the integrated graphical elements support interactivity that is not given by analog display holograms. This concept may enable new hybrid visualization schemes for areas where computer graphics alone does not yet provide the required realism, and computer-generated holography or electroholography does not yet offer the expected performance and quality. Assuming that Moore's law proves to hold in future, we can expect CGH playing a more dominant role for three-dimensional display technology.

REFERENCES

1. O. Bimber, Combining Optical Holograms with Interactive Computer Graphics, IEEE Computer, pp. 85-91, January 2004.
2. J.S. Kollin, S.A. Benton, and M.L. Jepsen, Real-Time Display of 3-D Computed Holograms by Scanning the Image of an Acousto-Optic Modulator, Proceedings of SPIE, vol. 1136, Holographic Optics II: Principles and Applications, Springer, pp. 178-185, 1989.
3. M. Lucente, Interactive Three-Dimensional Holographic Displays: Seeing the Future in Depth, ACM Computer Graphics, Special Issue on Current, New, and Emerging Display Systems, ACM Press, pp. 63-67, 1997.
4. M. Lucente and A. Tinsley, Rendering Interactive Images, Proceedings of ACM Siggraph'95, ACM Press, pp. 387-394, 1995.
5. M. Lucente, Interactive Computation of Holograms Using a Look-Up Table, Journal of Electronic Imaging, vol. 2, no. 1, pp. 28-34, 1993.
6. C. Petz and M. Magnor, Fast Hologram Synthesis for 3D Geometry Models Using Graphics Hardware, Proceedings of SPIE'03, Practical Holography XVII and Holographic Materials IX, vol. 5005, Springer, pp. 266-275, 2003.
7. F. Dreesen and G. von Bally, Color Holography in a Single Layer for Documentation and Analysis of Cultural Heritage, In Proceedings of Optics within Life Sciences (OWLS IV), Springer, pp. 79-82, 1997.
8. F. Dreesen, H. Deleré, and G. von Bally, High-Resolution Color Holography for Archaeological and Medical Applications, In Proceedings of Optics within Life Sciences (OWLS V), Springer, pp. 349-352, 2000.
9. D. Vila, Folding Time into Space – Novel Approaches to Integral Holography, Proceedings of SPIE'93, Practical Holography VII: Imaging and Materials, vol. 1914, Springer, pp. 230-235, 1993.
10. D. Vila, Holo-Dynamics – Linking Holography to Interactive Media, Proceedings of SPIE'93, Practical Holography VIII, vol. 2176, Springer, pp. 166-171, 1994.

11. M. Okamoto, H. Ueda, I. Nakamura, E. Shimizu, and T. Kubota, Multiple Illuminations Method for Rainbow Holograms using an LCD Projector, Proceedings of SPIE'97, Practical Holography XI, vol. 3011, Springer, pp. 70-75, 1997.
12. K. Yamasakia, M. Okamotoa, I. Nakamura, and E. Shimizu, Fluctuation Hologram with Multiple Images, Proceedings of SPIE'98, Practical Holography XII, vol. 3293, Springer, pp. 139-144, 1998.
13. J.R. Andrews and W.H. Haas, Compact Illuminators for Transmission Holograms, Proceedings of SPIE'89, Practical Holography III, vol. 1051, Springer, pp. 156-159, 1989.
14. J. Bongartz, D. Giel, P. Hering, Fast 3D topometry in medicine using pulsed holography, Proceedings of SPIE'02, Practical Holography XVI and Holographic Materials VIII, vol. 4659, 2002.
15. S. Frey, J. Bongartz, D. Giel, A. Thelen, and P. Hering, Ultrafast holographic technique for 3D in situ documentation of cultural heritage, Proceedings of SPIE'03, Optical Metrology for Arts and Multimedia, vol. 5146, pp. 194-201, 2003.
16. T. Carlsson, J. Gustafsson, and B. Nilsson, Development of a 3D camera, Proceedings of SPIE'99, Practical Holography XIII, vol. 3637, pp. 218-224, 1999.
17. N. Abramson, Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena, Applied Optics, vol. 22, no. 2, pp. 215-231, 1983.
18. E.B. Champagne, Nonparaxial Imaging, Magnification, and Aberration Properties in Holography, Journal of the Optical Society of America, vol. 57, no. 1, pp. 51-54, 1967.
19. K. Bazargan, Techniques in Display Holography, Ph.D. Dissertation, Physics Dept, Imperial College, London University, p. 16, 1986.
20. C. W. Slinger, C. D. Cameron, S. J. Coomber, R. J. Miller, D. A. Payne, A. P. Smith, M. A. G. Smith, and M. Stanley, Recent developments in computer generated holography: towards a practical electroholography system for interactive 3D visualization, Proceedings of SPIE'04, Practical Holography XVIII, vol. 5290 ,pp. 27-41, 2004.
21. C. W. Slinger, C. D. Cameron, and M. Stanley, Computer-generated holography as a generic display technology, IEEE Computer, vol 38, no. 8, pp. 46-53, 2005.
22. O. Bimber, T. Zeidler, A. Grundhöfer, G. Wetzstein, M. Möhring, S. Knödel, and U. Hahne, Interacting with Augmented Holograms, Proceedings of SPIE, Practical Holography XIX: Materials and Applications, Springer, January 2005.
23. S. Rusinkiewicz and M. Levoy, QSplat: A Multiresolution Point Rendering System for Large Meshes," Proceedings of ACM SIGGRAPH, pp. 343-352, 2000.
24. P.J. Besl, N. D. McKay, A Method for Registration of 3-D shapes, Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 14 no. 2, pp. 239-256, 1992.
25. Y. Gentet and P. Gentet, Ultimate emulsion and its applications: a laboratory-made silver halide emulsion of optimized quality for monochromatic pulsed and full-color holography, Proceedings of SPIE'00, Holography 2000, vol. 4149, pp. 56-62, 2000.
26. S.J. Hart and M.N. Dalton, Display holography for medical tomography, Proceedings of SPIE'90, Practical Holography IV, vol. 1212, pp. 116-135, 1990.

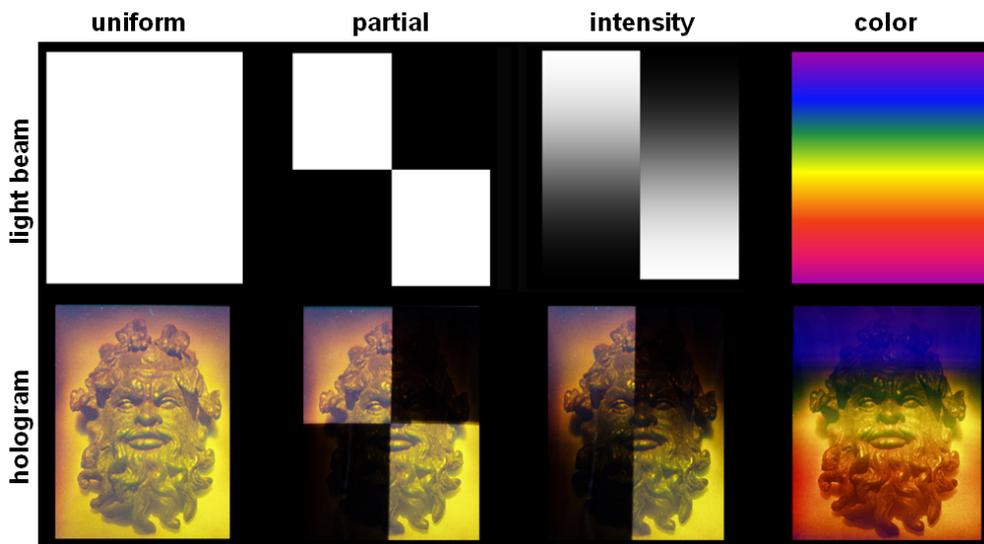


Figure 1: Replaying a monochrome white-light reflection hologram (bottom row) with a projected digital light beam (top row). Image reprinted from [22] – © SPIE.

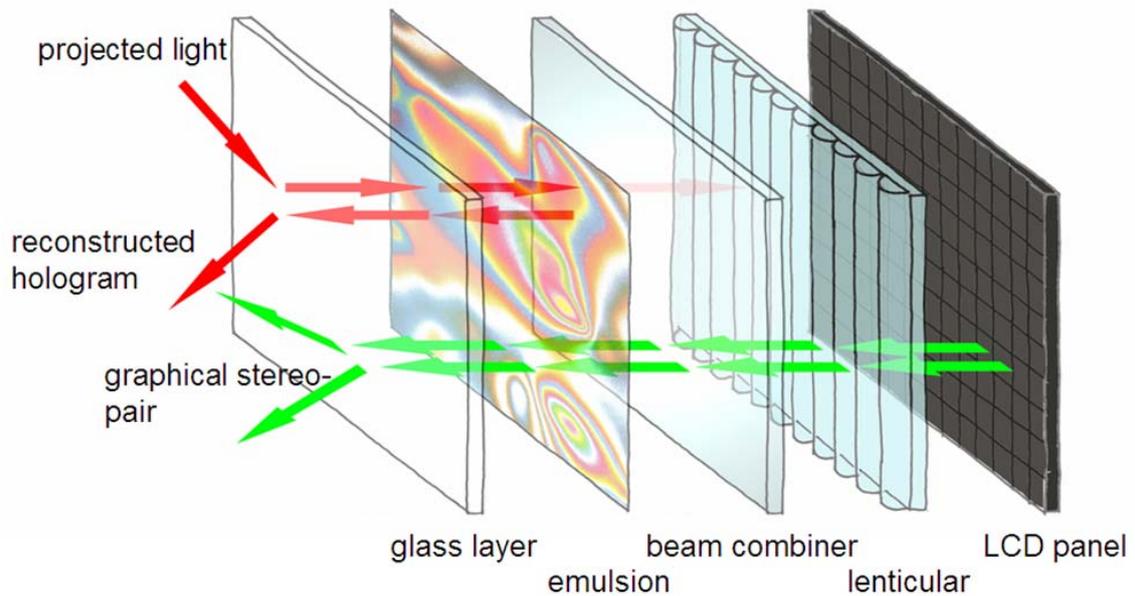


Figure 2: Explosion model of the optical layers' stacked structure. The example shows a transmission hologram in combination with autostereoscopic lenticular screen. Image reprinted from [1] – © IEEE.



Figure 3: A rainbow hologram of a dinosaur skull with autostereoscopically integrated graphical content of reconstructed soft-tissue and bones: consistent occlusion effects (top-left), projected illumination image blocked by a diffuse screen (top-right), consistent shading and shadow effects through synthetic re-illumination (bottom). Image reprinted from [1] – © IEEE.

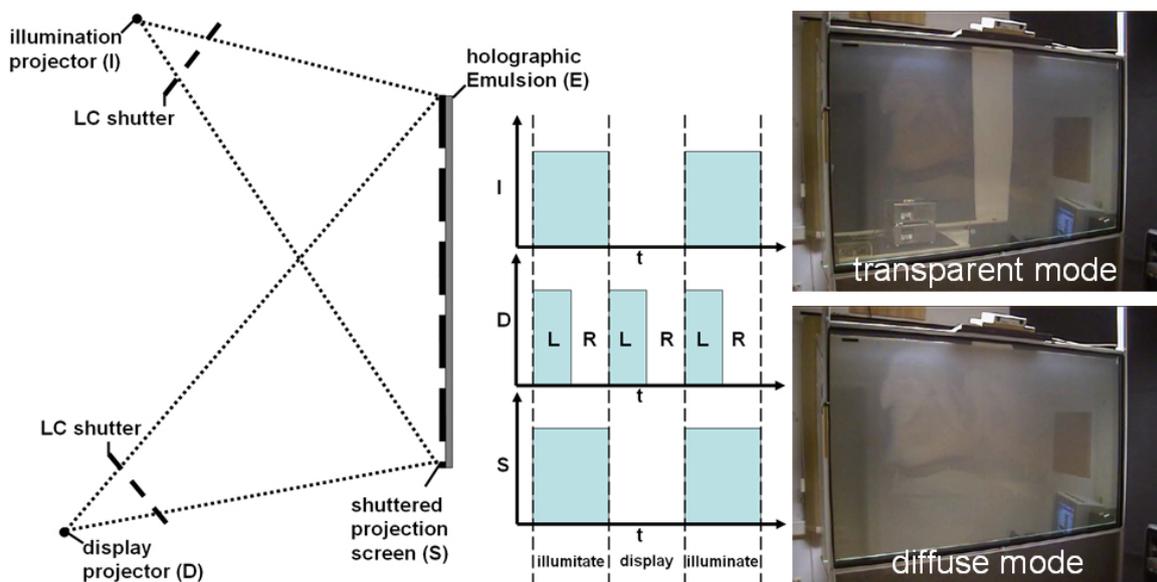


Figure 4: A phase dispersed liquid crystal screen (S) shuttered with 50Hz. In the transparent mode the illumination image is projected from I to replay the hologram. In the diffuse mode the left (L) and right (R) stereo images are projected from D to augment the hologram.

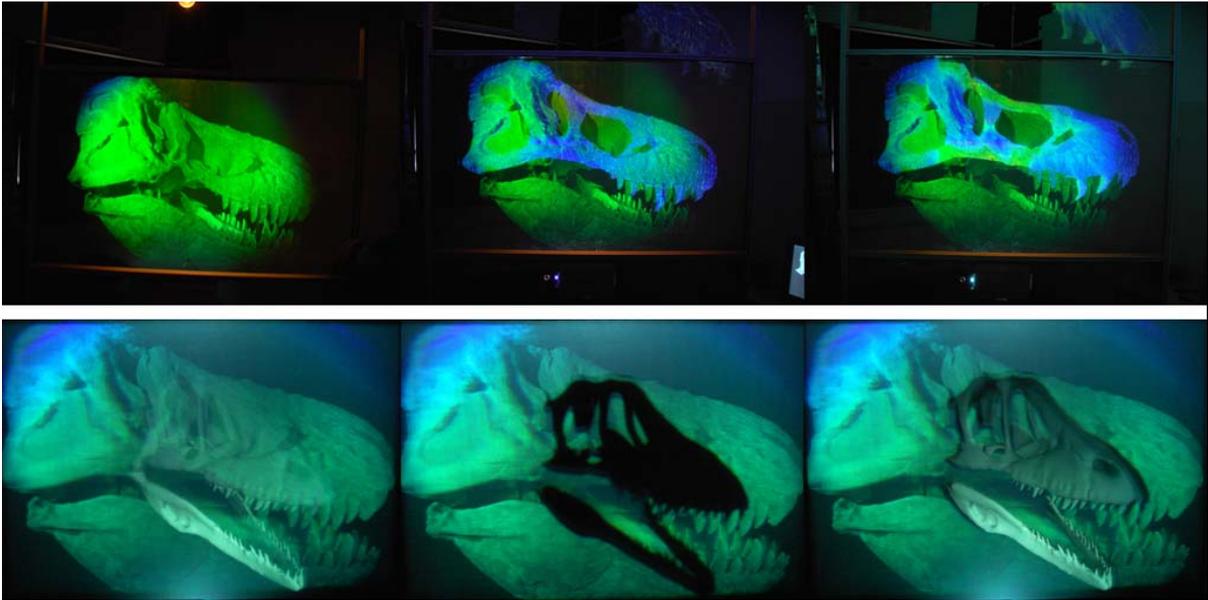


Figure 5: A large rainbow hologram of a T.Rex skull with integrated stereoscopic graphics: An animated FEM simulation of cranial mechanics and feeding projected directly onto the holographic film (top row). A model of a Deinonychus skull projected onto the synchronized PDLC screen (bottom row): without controlled illumination (bottom-left), without integrated graphics (bottom-center), with consistent occlusion effects (bottom-right).

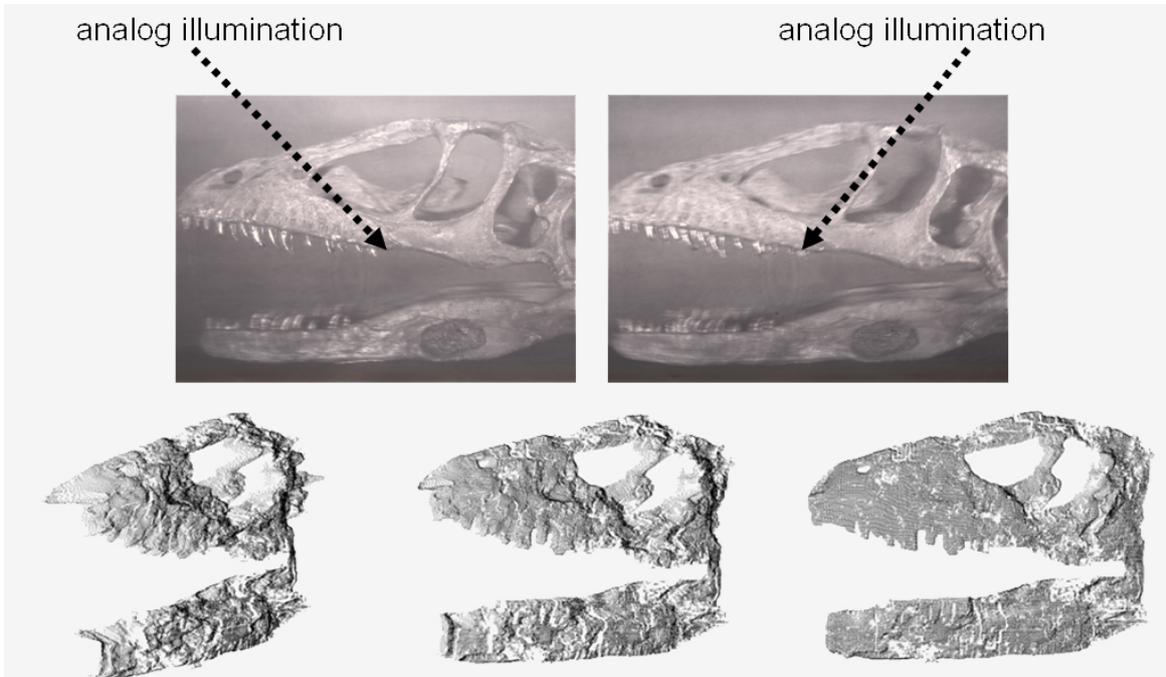


Figure 6: Two scans of a 40x30cm rainbow hologram with a flatbed scanner under different illumination angles (top), and the reconstructed depth of the recorded content (bottom). Image reprinted from [22] – © SPIE.

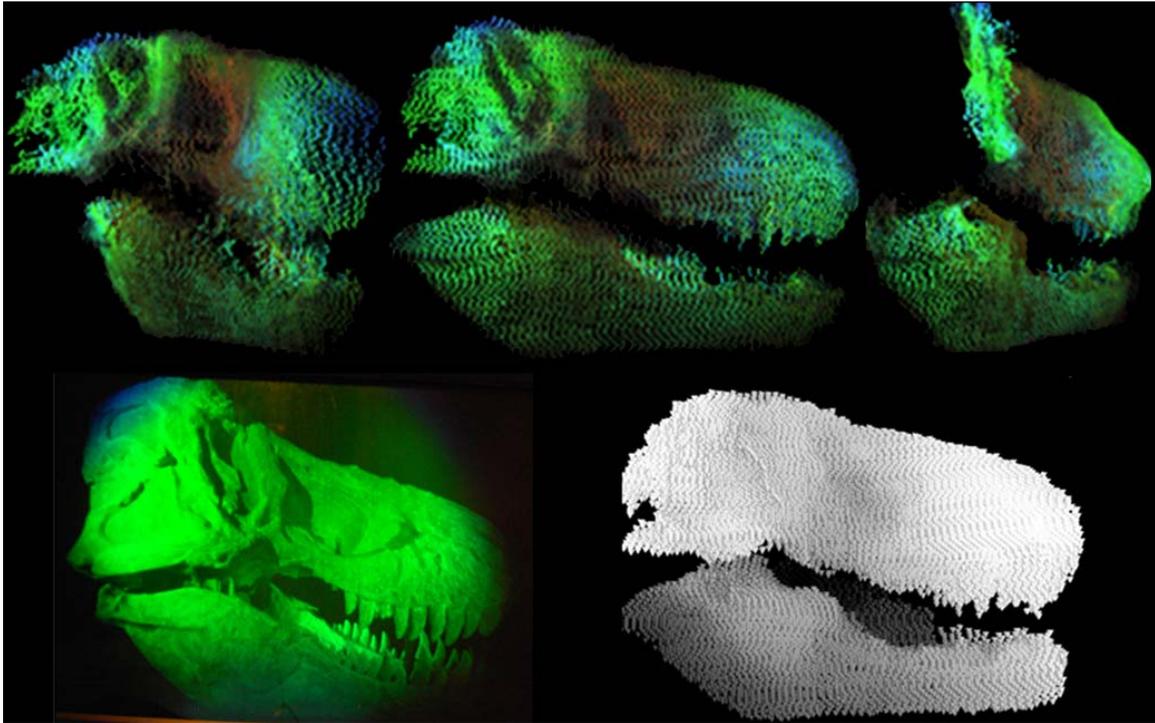


Figure 7: Reconstructed depth from 170cm x 103cm rainbow hologram (bottom-left): computed surface point cloud observed from different perspectives (top row), registered point clouds splatted into a depth map (bottom-right).

Digital Illumination for Augmented Studios

Oliver Bimber, Anselm Grundhöfer, Stefanie Zollmann and Daniel Kolster

Bauhaus-University Weimar
Bauhausstr. 11, 99423 Weimar
{bimber, grundhoefer, zollmann, kolster}@medien.uni-weimar.de
www.uni-weimar.de/medien/AR



Figure 1: Digital studio illumination: Synthetic illumination studio environment with projected spatial augmentations and video augmentations (left). Embedded imperceptible patterns for scene analysis and camera tracking (right).

Abstract

Virtual studio technology plays an important role for modern television productions. Blue-screen matting is a common technique for integrating real actors or moderators into computer generated sceneries. Augmented reality offers the possibility to mix real and virtual in a more general context. This article proposes a new technological approach for combining real studio content with computer-generated information. Digital light projection allows a controlled spatial, temporal, chrominance and luminance modulation of illumination – opening new possibilities for TV studios.

1 Introduction

Many modern TV productions apply virtual studio technology. A good overview can be found at [Gib98]. Chromakeying is the principle method for superimposing the live captured or recorded video signal of a physical blue screen studio with virtual content. Thereby, the video signal is analyzed and video pixels with a predefined color (e.g. blue or green) are replaced by computer-generated graphics. This allows using image processing techniques to efficiently separate the foreground from the background, and consequently to integrate real objects (such as an actor or a moderator) seamlessly into a purely virtual environment.

Blue screen techniques, however, limit virtual studio technology to special recording environments. Therefore, recent research initiatives investigate the potential of augmented reality (AR) for TV productions. In contrast to blue screen studios, fully equipped real television studios are

augmented with virtual content by superimposing the recorded video stream with computer graphics. According to virtual studios, we want to refer to this as *augmented studios*.

Several groups have already shown the advantages of augmented reality in a studio production context: Yama et al. [Yam02], for instance, augment 360° ultra high-definition omnidirectional images of artificial backgrounds –being distorted in real-time relative to the rotation of a pan-tilt camera, and being occluded by a real actor. An axi-vision camera [Kaw00] is used for simultaneously capturing color and depth information per pixel.

Recent examples are also shown in the context of the EU funded project MATRIS [Mat04]: Frahm et al. [Fra05] use a fish eye camera in addition to a studio camera. While the studio camera records the video content to be augmented, the fish eye camera observes the upper hemisphere to track the installed studio lights. Applying a structure from motion algorithm [Koc05] to both images makes the estimation of the studio camera's pose possible. Standard stereo algorithms [Koc98] allow reconstructing the depth of the studio setting and consequently enable correct occlusion effects between real and virtual objects. Furthermore, the knowledge of the real studio light sources allows computing a light map that ensures a consistent illumination and shadowing.

Virtual and augmented studio productions have to solve several technical challenges. One of them is the robust and fast *tracking* of the studio cameras [Ber03]. Some approaches apply special tracking hardware, while others try to estimate the

cameras' pose by observing natural features (e.g., ceiling-mounted studio lights or the studio content itself) or artificial tags [Tho97] with additional cameras, as explained above. Optical tracking is becoming more and more popular due to its robustness against most environmental disturbances, speed and precision. Optical tracking approaches can be categorized into *marker-less* and *marker-based* methods. Marker-less techniques, on the one hand, strongly rely on the robust detection of natural scene features [Fra05]. They will fail for uniformly structured surfaces or under dim lighting conditions. This limits the application of such techniques in TV studios to optimized situations. Marker-based tracking, on the other hand, provides artificial visual features by integrating detectable marker tags. However, these markers should neither be directly visible within the studio environment nor appear in the recorded video stream. Consequently, marker-based tracking is usually restricted to observing out-shot areas –such as the ceiling or the floor– which are normally covered by studio equipment, like light installations, cables, and mountings. Thus, occlusions and dynamic reconfigurations of the installations cause additional problems for marker-based tracking.

Another problem is the *acquisition of the scene depth*. This is necessary for integrating synthetic 3D objects into the video stream while producing consistent occlusion and illumination effects with the recorded real content. Some approaches reconstruct the scene geometry offline (during a special calibration step) using multi-viewpoint stereo from un-calibrated video sequences. The quality of such techniques relies on the quality of feature matching in the stereo pairs. However, finding matchable features to support a high quality depth reconstruction might be difficult – not only for real studio environments, but also for virtual studios or embedded blue screens that mainly apply uniformly colored matting surfaces. Besides offline reconstruction of the static studio setting, online depth estimations (e.g., of moving people in the scene) is even more problematic.

Yet another challenge for virtual and augmented studios is the question of how to *display direction information* to moderators, actors or participants during a live broadcast or a recording. Teleprompts or fixed screens offer limited possibilities since they do not allow bringing the presented information into a spatial context. Step sequences, for instance, are usually marked statically on the floor ground.

In this article, we propose the application of digital light projection for studio illumination – either exclusively or in combination with an existing analog lighting (cf. figure 1). This can solve several of the problems that are mentioned above, but also opens new possibilities for modern television productions.

The remainder of this article is organized as follows: Chapter 2 presents the technical key

concept of digital studio illumination, while chapter 3 presents early laboratory examples. Chapter 4 finally outlines open problems and challenges that have to be addressed when transferring the concept and the presented technological approaches to real studio environments.

2 Digital Studio Illumination

Projectors allow a spatial and temporal modulation of light that can be computer controlled and synchronized with the recording process of studio cameras. Our technical key concept is visualized in figure 2: Multiple projectors are used exclusively, or in combination with analog light sources for illuminating the entire, or parts of the studio environment.

Physically, projectors represent point light sources. Their capability of spatially modulating luminances and chrominances on a per-pixel basis, however, allows for computing and creating almost arbitrary shading effects within the studio synthetically. This is called *projector-based illumination* and is described in more detail in section 3.4. Compared to a conventional analog illumination, a projector-based illumination allows re-illuminating the studio on the fly – without changing the physical light sources. It can be combined with a similar technique that we refer to as *screen-based illumination*, which is also described in section 3.4.

Besides a spatial modulation, a temporal modulation of the projected light enables displaying different portions of the illumination time-sequentially. Figure 2 shows an example of two sequentially projected images (l and r) that carry different parts of the illumination. Due to the lack of the human visual system, the full illumination $l+r$ will be perceived when images are projected fast enough. The projection of more than two images is also possible, but requires fast display and capturing rates. This allows integrating coded patterns into one or several of these time slices in such a way that the sum of all images will result in the desired illumination or pictorial content. Synchronizing the studio cameras to the projection and capturing all slices separately, however, will make the coded patterns visible to the camera but not inside the studio. Summing all slice images computationally after recording will lead to the same image that would be integrated optically by the camera sensor during the shutter time that equals the duration required for projecting all slices (i.e., the fully illuminated studio). This is referred to as *embedded imperceptible pattern projection* and is explained in section 3.1. The extracted coded patterns can be used for camera pose estimation (section 3.2) or for depth reconstruction (section 3.3).

Embedded imperceptible pattern projection has recently been realized for 3D video capturing and continuous projector calibration applications: Cotting et al. [Cot04], for instance, synchronize a

camera to a well-selected time-slot during the modulation sequence of a DLP projector. Within this time-slot the calibration pattern is displayed and detected by the camera. Since such an approach requires modifying the original colors of the projected image, a loss in contrast is an undesired side effect. Other techniques rely on a fast projection of images that cannot be perceived by the observer. This makes it possible to embed calibration patterns in one frame and compensate them with the following frame. Capturing altering projections of colored structured light patterns and their complements allows the simultaneous acquisition of the scene's depth and texture without loss of image quality [Was05, Vie05]. Using structured light, depth reconstruction becomes easier since natural feature detection is not necessary because correspondences are provided by the projected codes. This also applies for camera tracking (section 3.2). Once the studio scene and the poses of the cameras are reconstructed, computer generated graphics can be augmented consistently into the recorded video stream.

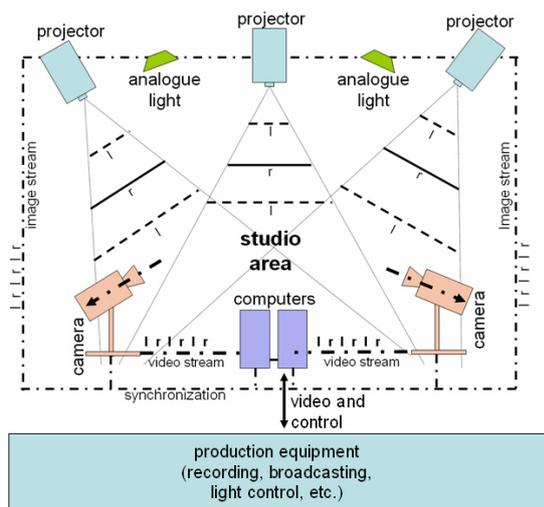


Figure 1: Modulating digital light in TV studios.

Furthermore, a temporal modulation of light supports the visualization of information that is visible in the studio, but not in the camera image. Fukaya et al. [Fuk03], call this *invisible light projection* with respect to invisibility to a camera rather than to an observer. They project an image onto a blue screen located in a real TV studio. The projected image is alternately blocked and transmitted by an LCD shutter mounted in front of the projector lens. A separate shutter control unit synchronizes projection and exposure time of a studio camera in such a way that images are only captured when the projection is blocked. Chromakeying can then be applied in a conventional way.

Shirai et al. [Shi05] apply chrominance and luminance keying instead of a shuttered projection for solving the same problem. Projecting an image onto a blue screen enables computing

both – a luminance key and a chrominance key that allow masking out the blue screen area for a final composition. This is possible only if the projected image is not brighter than the studio illumination.

Fukaya's simple concept can be extended to combine the presentation of arbitrary information that is visible only in the studio together with other data (e.g., the synthetic studio illumination) that is recorded simultaneously (section 3.1). This, for instance, makes it possible to display direction, moderation and other information dynamically and spatially anywhere within the studio – not being limited to fixed locations, like Teleprompts or static screens.

Problematic for most of the techniques describe above is the fact that the images are projected onto complex surfaces of a real studio rather than onto geometrically and radiometrically uniform surfaces, as it is the case for blue-screen studios. Projected images will be modulated (e.g., color blended or geometry distorted) by the underlying surfaces. Projected code patterns or direction information will not appear correct – neither in the studio nor in the captured images. Appropriate correction techniques have to be applied to compensate for these effects before the images are projected. Some of these techniques will be outlined in section 3.3.

It should be noted at this point that the projected images do not necessarily have to contain pure illumination information. They can carry an arbitrary content, such as projected pictures or special effects. Thus, it is interesting to investigate the combination of *projector-based spatial augmentation* [Bim05b] and conventional *video augmentation* for future virtual and augmented studio productions.

3 Bits and Pieces in the Small Scale

This chapter presents several proof-of-concept realizations for different technological components that have been mentioned in chapter 2. These components have not yet been transferred to a real studio environment but rather give an indication for the feasibility of our concept. Chapter 4 will discuss the remaining challenges when putting these pieces together in the large scale.

3.1 Embedding Imperceptible Patterns and Projecting Invisible Light

Digital Micro Mirror Devices (DMDs) that are used in DLP projectors control the pixel intensities by modulating the time in which the mirrors reflect light towards the projection surface. The micro mirrors can switch between their on/off states within approximately 10 μ s. Colors are modulated in addition by synchronizing a rotating color wheel to the DMD. Thus, the light projected per pixel by a DLP projector can be seen as a time-multiplexed sequence of a discrete number of color bands (e.g., for red, green and blue). Each color slot itself is intensity modulated by a sequence of on/off

states of the DMD that is encoded as a bit chain which represents the corresponding intensity value.

Imperceptible binary patterns can be embedded into a small time segment of this sequence. They are visible only to a camera which is synchronized to the same time segment. Cotting et al. [Cot04] occupies specific time slots exclusively for displaying a binary pattern. However, pixels of the original image that have colors and intensities which are modulated within these time slots cannot be displayed if their corresponding code pattern is turned off. They must be modified in such a way that they do not fall into these code slots which results in the reduction of tonal resolution of the projected image. Furthermore, the mirror flip sequences have to be measured precisely for each projector. Thus, an individual calibration of each projector is required. The advantage of this approach, however, is that it can be used together with off-the-shelf projectors.

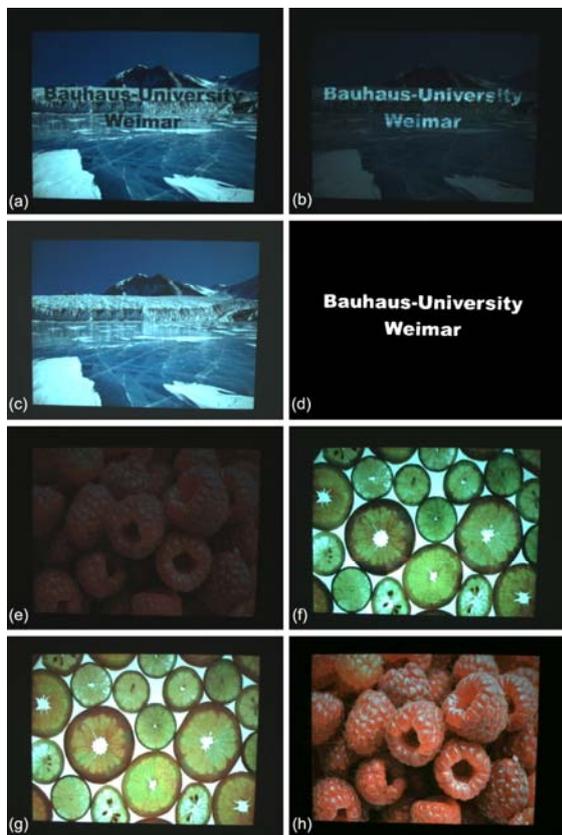


Figure 3: Embedded imperceptible patterns (a-d) and projected invisible light (e-h).

Instead of allocating time slots of a DLP projector's modulation sequence, code patterns can also be displayed within an entire projection frame. To ensure that they are not visible, the complementary pattern has to be projected immediately after. If this is being done fast enough, a white image will be perceived. A synchronized camera can capture both images – the code image and its complement. Adding both images computationally results in the image of the scene

illuminated with projected white light. This technique is used for capturing scene colors and geometry in the context of 3D video applications [Vie05, Was05]. One main drawback of this approach is that the code patterns are projected with the limited frame rate of the projector (e.g. 60Hz-70Hz for conventional projectors). This results in a well perceivable flickering and is consequently not appropriate for a real studio illumination.

In our approach, we combine the advantages of the techniques discussed above. This leads to a new method that embeds imperceptible patterns into projected images without perceivable flickering, reductions in contrast, or the need for a projector individual mirror flip calibration. Furthermore, it is neither limited to DLP technology, nor does it require extremely short exposure times of cameras. However, we do apply special stereoscopic projectors that are capable of a high native frame rate (120Hz in our case).

As discussed earlier, projecting two complementary images (l and r) with a high frequency makes them visually appear as the sum ($l+r$) of the two images (cf. figure 3c). A synchronized camera, however, can capture both images individually (cf. figure 3a,b). Subtracting both images ($l-r$) and binarizing the result allows extracting the embedded code (cf. figure 3d). Both images can also be added computationally to determine the image that is actually visible. This image can then be recorded (it equals figure 3c).

One challenge is to avoid visible flickering even if the code pattern is exchanged during the projecting sequence. In our solution, we smoothly fade new code patterns in and out within a short sequence of subsequent projection frames. Dynamic content, such as videos or interactive applications have to be synchronized to this process to ensure that corresponding frames contain the same visible content. Currently, we achieve a frame rate of 20Hz for displaying dynamic content and simultaneous code extraction (while projecting with 120Hz). This speed is mainly limited by the capturing rate of our camera¹. By applying a faster camera² we estimate to double this frame rate approximately.

As mentioned in chapter 2, it is also possible to display information that is visible in the studio but is not recorded by the camera. An example is illustrated in figure 3. Assuming a desired visible studio image v , we know that it will be composed from two or more sequentially projected slice images (e.g., $v=l+r$ for two slices). Knowing the image that should be captured by the camera in addition (e.g., l), we have to compute a compensation image r in such a way that $r=v-l$, $v>=l$. Figure 3e illustrates the image l that is only visible to the camera when capturing it at the corresponding time slot. Figure 3f presents the

¹ A Point Grey Dragonfly 2.

² Such as a Dragonfly Express.

computed compensation image r for the desired image v that is visible as projection in the studio (cf. figure 3g). Figure 3h illustrates a digitally contrast enhanced version of I . This image is actually being recorded simultaneously (instead of the low contrast version shown in figure 3e).

Note that the methods described in this section are widely independent of the image content. Thus, the projected images do not necessarily have to contain pictorial data as in the example shown with figure 3. They can also carry synthetic illumination information instead. This will be described in section 3.4.

3.2 Camera Pose Estimation with Embedded Markers

Embedding imperceptible patterns into projected images that carry pictorial or illumination information allows supporting a marker-based in-shot camera tracking. This means that markers can be displayed directly within the studio environment. However, they are not directly visible within the studio itself and will neither be recorded. In contrast to out-shot approaches (e.g., when tracking markers or studio lights that are attached to the ceiling), this does not cause conflicts with other studio equipment and does not require an additional cameras or other devices.

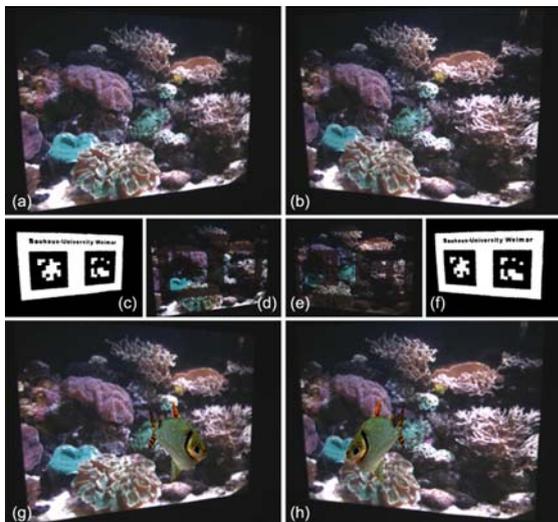


Figure 4: Embedded imperceptible markers for in-shot camera tracking and video augmentation.

Figures 4a and 4b show a projected image as it can be perceived when observing it from two different perspectives. The projected image contains an embedded imperceptible marker that is extracted with the technique explained in section 3.1. This is illustrated in figures 4c-f for both perspectives. After binarizing the marker the position and orientation of the camera relative to the marker's origin can be determined³. The two captured images I and r that are required for

separating the embedded code are computationally added. A virtual object (the fish) is then rendered on top of the combined image from the estimated camera perspective (cf. figures 4g,h). This leads to a live video augmentation of the captured environment that contains arbitrary projections. The augmentation, however, will not appear in the real environment.

Once again it has to be noted that the projected images do not necessarily have to contain pictorial information, as shown in figure 4. They can carry the studio illumination instead (see section 3.4) – making it possible to augment the illuminated studio environment with exactly the same method.

3.3 Projecting onto Everyday Surfaces and Acquisition of Scene Depth

The application of projection technology in virtual or real studios is usually restricted to the projection of images onto special screen surfaces, such as white diffuse screens or blue screens. The concept that is proposed in this article, however, requires a projection onto the real surfaces of the existing studio setting – and possibly onto dynamic content, such as moderators and actors. As discussed above, these images can carry pictorial information – making it possible to display visible direction information dynamically onto arbitrary studio surfaces without being limited to static screens. The image can display imperceptible embedded patterns that are used for camera tracking, depth acquisition, or online calibration processes. Finally, they can also contain the synthetic studio illumination (see section 3.4). In any case, the projected images are distorted when being reflected from non-optimized surfaces. This does not only lead to visible errors in projected pictorial content which are well perceivable in the studio or in the recorded video stream, but also to problems when extracting embedded code patterns. Thus, real-time image correction techniques are required that are capable of compensating for any image distortion that is caused by a projection onto arbitrary surfaces. Furthermore, multiple projectors have to be calibrated in such a way that a single consistent image can be presented from multiple individual projector contributions.

Geometric calibration techniques for multi-projector systems, such as tiled screens, widely use camera feedback to support automatic registration. Geometric image registration approaches for simple planar surfaces determine homography matrices when warping images from a reference perspective to the perspectives of the projectors [Yan01]. For projection surfaces with a non-trivial but known geometry, intrinsic and extrinsic projector parameters have to be estimated to enable image warping via projective texture mapping operations [Ras99].

These conventional techniques alone will fail in the case of displaying images with projectors

³ ARTag [Fia05] was used for marker tracking.

in a complex studio environment, because the surfaces available in a real studio are usually not optimized for projections. They can be geometrically complex, colored, textured, non-diffuse, and can cover a large depth range. This results in geometric distortions, color blending, intensity variations, and regional defocus effects in the projected images.

With coded structured light techniques, the surface geometry, reflectance as well as the global and local illumination behavior can be automatically determined by evaluating the corresponding camera feedback. As explained above, the structured code can be embedded seamlessly into the projected content and remains imperceptible in the studio or in the recorded content.

This makes it possible to determine the relation between all pixels of each projector with respect to the parameters of the scene points onto which they project (i.e., their geometric position, reflectance, local and global illumination parameters).

Knowing these parameters, many distortion effects can be compensated on a per-pixel basis and in real-time when implementing on modern GPUs. Besides pixel-precise geometric warping [Bim05a], photometric [Bro05] and radiometric calibration [Nay03, Gro04, Bim05a, Fuj05, Ash06, Gru06, Wet06] techniques ensure chrominance and luminance consistency, as well as the compensation of color and intensity artifacts when projecting onto colored and textured surfaces. Multi-focal projection [Bim06b] and defocus compensation [Zha06] methods can be used for increasing the focal depth of single or multi-projector systems. Reflection highlights on specular surfaces can also be eliminated with appropriate multi-projector techniques [Par05]. Even global illumination effects, such as surface-to-surface scattering [Bim06a] or complex physical light modulations like inter-reflections and refractions [Wet06] can be neutralized.

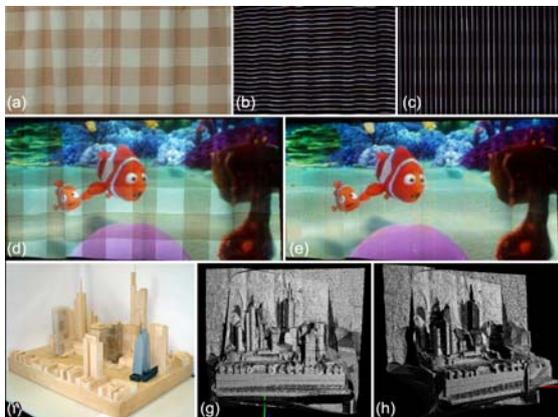


Figure 5: Projection onto a colored window curtain with and without radiometric compensation (a-e), ©2005 IEEE. Acquired scene depth via structured light projection (f-h).

An example for a projection onto a non-optimized surface is illustrated in figures 5a-e. The pixels of an uncompensated image projected onto a colored and textured surface will be color and intensity blended with the underlying surface pigments (cf. figures 5a,d). After measuring the surface parameters via structured light projection (cf. figures 5b,c), a real-time and pixel-precise radiometric compensation can be applied. This minimizes these artifacts directly on the surfaces and consequently in the captured images (cf. figure 5e). Note that in both cases the image is warped (also in real-time and on a per-pixel basis) to compensate for geometric distortions caused by the non-planar surface. Thus, it appears as being projected onto a planar surface from the perspective of the camera.

All of these image correction techniques ensure that the desired images can be displayed and captured when projecting synthetic illuminations, pictorial content or code patterns directly onto real studio surfaces. The parameters that are required for carrying out the compensation computations can be determined continuously and unnoticeably with the techniques described in section 3.1. This, for example, implies that the coded patterns shown in figures 5b and 5c are not visible in the studio or in the recorded video. Besides radiometric and geometric distortions, other effects, such as global inter-reflections, specular highlights or defocus effects can be compensated as described earlier. To explain these techniques is out of the scope of this article.

Furthermore, being able to project corrected code patterns onto arbitrary surfaces allows applying structured light projection for a fast depth acquisition more efficiently. This is shown in figures 5f-h. The depth map has been computed for a number of unstructured positions of a marker-tracked camera. Line-strip patterns (such as the ones shown in figures 5b,c) have been projected onto the scene for providing sufficient artificial features.

3.4 Projector-Based and Screen-Based Illumination

Synthetic re-illumination has been an active topic in computer graphics and computer vision for many years. We can differentiate between methods that re-illuminate recorded image or video content, or approaches that physically re-illuminate a real scene or object with controlled lighting. Only the latter category is of interest for our concept.

A technique called *virtual re-illumination* has been introduced by Paul Debevec [Deb02]. It was used to create special effects in recent Hollywood movies, such as Spiderman or King Kong. A special recording environment, called *LightStage*, allows producing a variety of different basic lighting situations in a high speed with analog

light bulbs surrounding an actor. They are captured with a synchronized fixed camera. Having recorded the discrete basis functions of a 5D slice of a 12D reflectance field, the illumination in the video content can be altered after recording.

Other recording environments surround actors with diffuse rear-projection screens instead of light bulbs. This makes a direct re-illumination before or during recording possible by displaying appropriate environment maps on the screens [Mit05]. We want to refer to this as *screen-based illumination*.

Both approaches, virtual and screen-based re-illumination, require specialized recording environments and are –so far– not suitable for real studios. They also focus on the re-illumination of actors or moderators, rather than on the re-illumination of an entire studio environment.

Projector-based illumination techniques re-illuminate a physical environment synthetically (before or during recording) using a discrete number of unstructured aligned projectors. Thereby, the projectors illuminate the real environment directly and on a per-pixel basis, rather than indirectly by projecting onto diffuse screens that scatter light into the environment.

The challenge of projector-based illumination is to produce a defined lighting situation within a real environment without the necessary light sources. The only available light sources are the projectors themselves that represent static point lights. Consequently, images have to be computed for each projector with the following objectives: First, they must neutralize the physical illumination effects that are caused by each projector as a real point-light source. Second, they have to produce the defined virtual lighting situation synthetically.

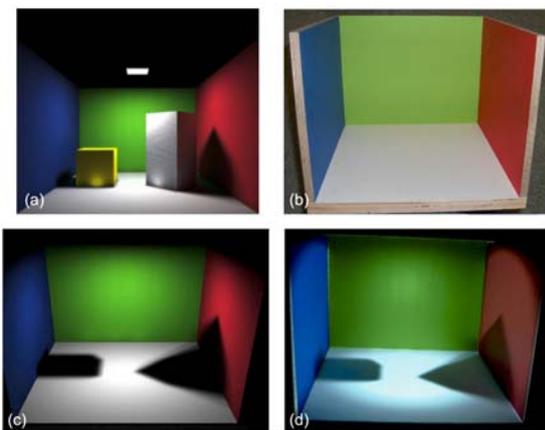


Figure 6: Projector-based illumination for creating global illumination effects synthetically, ©2003 IEEE.

Projector-based illumination has been demonstrated on the small scale: for uniform white surfaces simulating only local lighting effects [Ras01], for colored and textured surfaces

simulating local and global lighting effects [Bim03], and even for optical holograms [Bim04]. An image-based technique to relight real objects illuminated by a 4D incident light field, representing the illumination of an environment, has also been described by Masselus et al. [Mas03].

Figure 6 demonstrates an example for projector-based illumination. The global illumination for a model of a simple room environment, consisting of walls, boxes and an area light source is computed. Figure 6a shows a screenshot of the results after several radiosity iterations. Shading, soft-shadows and color-bleedings caused by the area light source and the model's geometry are clearly visible. A real miniature mock-up of the room without the boxes is illustrated in figure 6b. Using projector-based illumination, the computed lighting on the wall surfaces can be created syntactically. While figure 6c shows another screenshot of the computationally illuminated scene without boxes, figure 6d presents a photograph of the real mock-up from the same perspective, but illuminated synthetically with a single projector. Similar shading effects that are computed for the virtual area light can be produced by the projector (i.e., a physical point light being located at a completely different position). Minor differences between the computed illumination (figure 6c) and the synthetically created illumination in the real mock-up (figure 6d) are due to technical limitations of the projector, such as its high black-level and its low contrast.

The rendering of the intensity images that create the synthetic illumination is fully compatible with the techniques presented in section 3.1 and 3.2. The only difference is that projected images do not carry pictorial content only, but also the synthetic studio illumination.

Note that the defined virtual illumination can be arbitrary. Besides global and view-independent effects, such as the ones shown in figure 6, it can also contain view-dependent effects, such as synthetic specular highlights or refractions. The latter case requires the knowledge of the camera's poses. This can be determined as described in section 3.2.

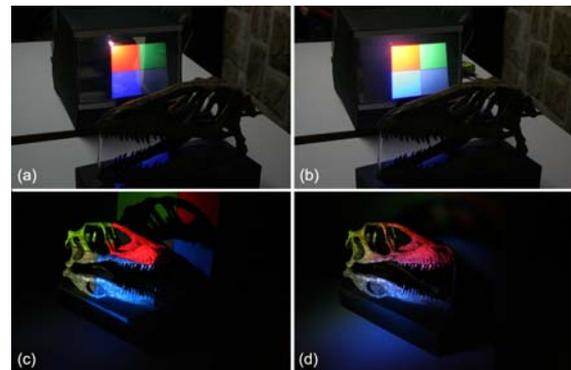


Figure 7: Combination of projector-based illumination (a,c) and screen-based illumination (b,d).

However, for a projector-based illumination it is essential that physical light modulations within the real environment are canceled out. This was discussed in section 3.3.

Projector-based and screen-based illumination can also be combined by projecting through a shuttered diffuse screen (cf. figure 7). Such screens contain a phase dispersed liquid crystal layer that can be electronically switched to a diffuse or a transparent state. In the transparent state images are projected straight through the screen (cf. figures 7a,c), and projector-based illumination or displaying pictorial content is supported. In the diffuse mode, the projected image is scattered at the screen (cf. figures 7b,d), and a screen-based illumination in a real recording environment becomes possible. The sizes of the diffusers affect the possible illuminating effects. They can reach from small screens that are rigidly mounted in front of each projector lens (as shown in figure 7) to large screens, as used for existing recording environments that currently apply passive projection screens [Mit05].

Since the screens can be switched between transparent and diffuse modes with a high speed (e.g., 50Hz-100Hz), a simultaneous activation of both illumination schemes can be perceived and recorded. Synchronizing the camera to the shutter signal of the screen allows separating both illumination types and combining them computationally, if necessary. If analog light sources are used in addition, they have to be shuttered and synchronized with projectors, screens and cameras.

For projector-based and screen-based illumination, an adequate number of projectors is required to lighten the entire studio environment from multiple directions. Multi-projector techniques, such as the ones outlined in section 3.3 can be used for calibration and for displaying consistent and undistorted images. Multi-projector techniques can also be used for the removal of shadows that are cast by static or dynamic content [Suk01, Jay01].

4 The Big Picture

With the concept proposed in this article, we envision a technical extension to existing virtual and augmented studio technology that enables new effects and control possibilities.

Modulated digital light projection opens new possibilities for modern television studios:

- dynamic re-illumination of studio settings, moderators and actors without physical modification of the lighting equipment;
- marker-based in-shot tracking of studio cameras without visible markers;
- dynamic presentation of un-recorded direction, moderation and other

information spatially anywhere within the studio;

- integration of imperceptible coded patterns that support continuous online-calibration, camera tracking, and acquisition of scene depth.

Most of these points can be addressed individually with different technological approaches. Digital light projection, however, holds the potential of offering a unified solution.

Although the examples presented in chapter 3 prove the feasibility of individual techniques in the small scale and under laboratory conditions, several problems have to be addressed before they can be applied in real studio environments. Up-scaling these techniques clearly represents the main challenge.

Installing a large number of projectors in a studio makes the investigation of a robust hardware infrastructure necessary. This includes adequate cooling for ceiling mounted projectors and analog lighting equipment, networked PCs for distributed rendering, synchronization electronics for projectors, cameras and shutter screens, and interfaces to existing production equipment (cf. figure 2).

Besides the hardware infrastructure, a scalable software framework has to be implemented that supports the following points:

- driving multi-projector cluster over a grid of networked PCs (including calibration);
- supporting fast frame grabbing and video processing;
- realizing real-time rendering techniques for multi-projector systems, such as synthetic re-lighting and image compensation;
- offering computer vision techniques for projector-camera systems, such as code extraction, pose-estimation, depth acquisition, and scene analysis;
- implementing consistent video and projector-based spatial augmentation techniques;
- interfacing to existing production software and equipment.

Automatic calibration methods will register the projectors to the studio environment – continuously if embedded into online presentations. Once the projectors are registered, imperceptible structured light projection can be used together with the tracked studio cameras to acquire scene depth of static (e.g., furniture) as well as dynamic (e.g., moving people) content. Techniques have to be developed that automatically extract the dynamic from the static content. While the static content can be scanned during an offline stage, dynamic components have to be scanned online. The

reconstructed depth information is required for creating consistent occlusion and illumination effects for graphical augmentations that are integrated into the recorded video stream. They are also essential for the synthetic re-illumination process.

Having geometric information of the static studio content allows analyzing its topology. This, for instance, enables the search for planar sub-surfaces that are suitable for displaying projected tracking markers. Doing this continuously makes it possible to dynamically reconfigure marker positions on the fly to ensure an optimal visibility and to avoid occlusions. This is clearly not possible with static markers [Tho97].

The development of scalable re-illumination techniques for creating synthetic illumination effects is as challenging and exciting as realizing scalable image compensation methods for neutralizing the physical modulation of light in real environments, such as TV studios. Furthermore, the investigation of efficient mixtures between video augmentations and projector-based spatial augmentations in a studio production context is yet another interesting topic of research.

Although this article focuses on spatial and temporal light modulation in TV studios, alternative modulated methods, such as wavelength modulation (e.g., on an invisible infrared basis), will also have to be investigated. Finally, we believe that these techniques will not only find their applications in television studios only. They will offer new possibilities to similar domains, such as photo studios (in the context of computational photography [IEE06]) or live stage performances.

Our group at the Bauhaus-University Weimar is actively working on realizing this concept and transferring our current small-scale implementations to the large scale.

Remark

The investigation and realization of concepts and techniques outlined in this article have been proposed to the Deutsche Forschungsgemeinschaft (DFG).

References

- [Ber03] Bernas, M., Teisler, M. Determination of the camera position in virtual studio Signal Processing and Information Technology. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on 14-17 Dec. pp. 535 – 538, 2003.
- [Gib98] Gibbs, S., Arapis, C., Breiteneder, C. Lalioti, V. Mostafawy, S. Speier, J. Virtual studios: an overview. IEEE Multimedia. vol.5, no.1, pp. 18-35, 1998.
- [Yam02] Yamanouchi, Y. et al. Real space-based virtual studio seamless synthesis of a real set image with a virtual set image. Proceedings of the ACM symposium on Virtual reality software and technology. Hong Kong, China. SESSION: Hybrid VR. pp. 194-200, 2002.
- [Kaw00] Kawakita, M. et al. 2000. Axi-vision camera: a three-dimension camera, In Three-Dimensional Image Capture and

Applications III. Proceedings of SPIE, Vol.3958, pp. 61-70, Jan. 2000.

[Tho97] Thomas, G.A. Jin, J. Urquhart, C. A versatile camera position measurement system for virtual reality tv production. International Broadcasting convention. Amsterdam. 12-16 September, pp. 284-289, 1997.

[Koc05] Markerless Image-based 3D Tracking for Real-time Augmented Reality Applications. R.Koch, K.Koeser, B.Streckel, J.-F. Evers-Senne WIAMIS 2005, April 13-15, Montreux, Switzerland, 2005.

[Fra05] Jan-Michael Frahm, Kevin Koeser, Daniel Grest, Reinhard Koch: Markerless Augmented Reality with Light Source Estimation for Direct Illumination. CVMP, London, December 2005.

[Koc98] R. Koch, M. Pollefeys, and L. Van Gool. Multi viewpoint stereo from uncalibrated video sequences. In Proceedings of ECCV'98, vol. 1 of LNCS, pp. 55-71, Freiburg, Springer-Verlag, 1998.

[Co204] Cotting, D., Naef, M., Gross, M., Fuchs, H.: Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display, IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR04), pp. 100-109, Arlington, 2004.

[Vie05] Vieira, M. B., Velho, L., Sá, A., Carvalho, P., C.: A Camera-Projector System for Real-Time 3D Video. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 96, Volume 3, San Diego, 2005.

[Was05] Waschbüsch, M., Würmlin, S., Cotting, D., Sadlo, F., Gross, M.: Scalable 3D Video of Dynamic Scenes. The Visual Computer, Volume 21, pp. 629-638, 2005.

[Mat04] Markerless real-time Tracking for Augmented Reality Image Synthesis, <http://www.ist-matris.org/>, 2004.

[Yan01] Yang, R., Gotz, D., Hensley, J., Towles, H.: PixelFlex: A Reconfigurable Multi-Projector Display System. IEEE Visualizations (IEEE Viz'01), pp. 161-168, 2001.

[Ras99] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., & Fuchs, H.: Multi-projector displays using camera-based registration, Proc. of IEEE Visualization (IEEE Viz'99), pp. 161-168, 1999.

[Bim05a] Bimber, O., Emmerling, A., Klemmer, T.: Embedded Entertainment with Smart Projectors. IEEE Computer, 38(1), pp. 56-63, 2005.

[Bro05] Brown, M., Majumder, A., and Yang, R., Camera-Based Calibration Techniques for Seamless Multi-Projector Displays. IEEE Transactions on Visualization and Computer Graphics, 11(2), pp. 193-206, 2005.

[Fuj05] Fujii, K., Grossberg, M.D., & Nayar, S.K. (2005). A projector-camera system with real-time photometric adaptation for dynamic environments. Proc. of Computer Vision and Pattern Recognition, 2, pp. 20-25, 2005.

[Gro04] Grossberg, M.D., Peri, H., Nayar, S.K., & Bulhumeur, P. (2004). Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 1, pp. 452-459, 2004.

[Nay03] Nayar, S.K., Peri, H., Grossberg, M.D., & Belhumeur, P.N. (2003). A Projection System with Radiometric Compensation for Screen Imperfections. Proc. of International Workshop on Projector-Camera Systems, 2003.

[Deb02] Debevec, P. Wenger, A. Tchou, C. Gardner, A. Waese, J. Hawkins, T., A lighting reproduction approach to live-action

- compositing. *ACM Transactions on Graphics (TOG)*, Proceedings of the 29th annual conference on computer graphics and interactive techniques SIGGRAPH '02: 21 (3), 2002.
- [Ras01] Raskar, R., Welch, G., Low, K.L. and Bandyopadhyay, D., Shader Lamps: Animating real objects with image-based illumination. *Proc. of Eurographics Rendering Workshop*, pp. 89-102, 2001.
- [Bim03] Bimber, O. Grundhöfer, A., Wetzstein, G., and Knödel, S., Consistent Illumination within Optical See-Through Augmented Environments. *Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'03)*, pp. 198-207, 2003.
- [Bim04] Bimber, O., Combining Optical Holograms with Interactive Computer Graphics. *IEEE Computer*, 37(1), pp. 85-91, 2004.
- [Fuk03] Fukaya, T., Fujikake, H., Yamanouchi, Y., Mitsumine, H., Yagi, N., Inoue, S., Kikuchi, H. An effective interaction tool for performance in the virtual studio – invisible light projection system. *NHK Science & Technical Research Laboratories, Japan*, 2003.
- [Shi05] Shirai, A., Takahashi, M., Kobayashi, K., Mitsumine, H., and Richir, S., Lumina Studio: Supportive Information Display for Virtual Studio Environments Workshop 03: Emerging Display Technologies, *IEEE VR 2005*.
- [Ash06] Ashdown, M., Okabe, T., Sato, I., and Sato, Y., Robust Content-Dependent Photometric Projector Compensation, *IEEE International Workshop on Projector-Camera Systems (PROCAMS)*, 2006.
- [Gru06] Grundhoefer, A. and Bimber, O., Real-Time Adaptive Radiometric Compensation. *ACM SIGGRAPH'06 (Poster)*, submitted to *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2006.
- [Wet0606] Wetzstein, G. and Bimber, O., Radiometric Compensation of Global Illumination Effects with Projector-Camera Systems. *ACM SIGGRAPH (Poster)*, submitted to *IEEE Transactions on Graphics (TOG)*, 2006.
- [Bim06b] Bimber, O. and Emmerling, A., Multi-Focal Projection: A Multi-Projector Technique for Increasing Focal Depth. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*. vol. 12, no. 4, pp. 658-667, 2006.
- [Zha06] Zhang, L. and Nayar, S. K., Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, pp. 907-915, 2006.
- [Bim06a] Bimber, O., Grundhoefer, A., Zeidler, T., Danch, D., and Kapakos, P. 2006. Compensating Indirect Scattering for Immersive and Semi-Immersive Projection Displays. *Proc. of IEEE Virtual Reality (IEEE VR'06)*, pp. 151-158, 2006.
- [Par05] Park, H., Lee, M.-H., Kim, S.-J., and Park, J.-I., Specularity-free projection on nonplanar surface. *Proc. of Pacific-Rim Conference on Multimedia (PCM)*, pp. 606-616, 2005.
- [Bim05b] Bimber, O. and Raskar, R., *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters LTD (publisher), ISBN: 1-56881-230-2, 2005.
- [Suk01] Sukthankar, S., Cham, T.J., and Sukthankar, G., Dynamic Shadow Elimination for Multi-Projector Displays. *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001.
- [Jay01] Jaynes, C., Webb, S., Steele, R., Brown, M. and Seales, W., Dynamic shadow removal from front projection displays, *Proc. of IEEE Visualization*, 2001.
- [Mas03] Masselus, V., Peers, P., Dutri, P., and Willems, Y. D., Relighting with 4D Incident Light Fields. *ACM Trans. Graph.* 22, 3, pp. 613-620, 2003.
- [Mit05] Mitsumine, H., Fukaya, T., Komiyama, S., Yamanouchi, Y., *Immersive Virtual Studio*, *ACM SIGGRAPH'05 Sketch*, 2005.
- [IEE06] *IEEE Computer, Special Issue on Computational Photography*, Levoy, M., Nayar, S., Debevec, P., Cohen, M., and Szeliski, R., (guest authors), Bimber, O. (guest editor), August issue, 2006.
- [Fia05] Fiala, M., ARTag, a Fiducial Marker System Using Digital Techniques, In proceedings of Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 590 – 596, 2005.

Real-Time Adaptive Radiometric Compensation

Anselm Grundhöfer and Oliver Bimber, *Member, IEEE*,

Abstract—Recent radiometric compensation techniques make it possible to project images onto colored and textured surfaces. This is realized with projector-camera systems by scanning the projection surface on a per-pixel basis. With the captured information, a compensation image is calculated that neutralizes geometric distortions and color blending caused by the underlying surface. As a result, the brightness and the contrast of the input image is reduced compared to a conventional projection onto a white canvas. If the input image is not manipulated in its intensities, the compensation image can contain values that are outside the dynamic range of the projector. They will lead to clipping errors and to visible artifacts on the surface. In this article, we present a novel algorithm that dynamically adjusts the content of the input images before radiometric compensation is carried out. This reduces the perceived visual artifacts while simultaneously preserving a maximum of luminance and contrast. The algorithm is implemented entirely on the GPU and is the first of its kind to run in real-time.

Index Terms—Computer graphics, picture/image generation, display algorithms, image processing, computer vision, radiometry, reflectance, enhancement, color

I. INTRODUCTION

Compact pocket projectors that are running from battery and communicate to laptops or cell phones via WiFi will support a maximum level of mobility in future. There is no doubt that LED technology will become bright enough to keep up with today's projector lamps. Yet one question still remains: On what to project on if not carrying around a projection canvas?

Projector-camera systems have been used together with radiometric compensation algorithms for projecting onto complex everyday surfaces, like papered walls or structured table tops. Most of the existing approaches consider only the characteristics of the surface - such as its reflectance or geometry - for the compensation of visual artifacts. The properties of the image to be displayed - like its brightness and contrast - are normally not taken into account. This can lead to clipping errors and to remaining visual artifacts at dark surface pigments that are due to the limited dynamic range and brightness of projectors. An adjustment of the image intensity to avoid this problem can only be carried out manually. Only recent approaches adapt the image content (sometimes based on the capabilities of the human visual system) for minimizing these artifacts [21] [1]. But because of their complex image transformations and numerical minimizations such algorithms are too complex to support real-time applications. Thus, animated or interactive content cannot be compensated in an optimized way. This, however, will be essential for future mobile projection systems.

We propose a novel algorithm that adjusts the image content to reduce visible artifacts in real-time. This is achieved by an analysis of the projection surface and the image content,

followed by a manipulation of the image's local and global intensity values before radiometric compensation is carried out. The result is a significant reduction of clipping errors and visible artifacts while preserving a high contrast ratio and brightness. The objective enhancement of the perceived visual quality for projected animated content is validated by an informal user study. The real-time capability of our algorithm enables an adaptive radiometric compensation for presentations of interactive and animated imagery onto non-optimized surfaces with consumer projectors.

A. Outline of the Article

The remainder of the article is organized as follows: We start with a discussion on relevant related work in section II, followed by an overview of our approach in section III. The algorithm is then described in detail: While section IV outlines the image and surface analysis components, section V describes the real-time adaptation and compensation technique. Examples are shown in section VI and a performance analysis of our algorithm is carried out in section VII. The results of an informal user study are presented in section VIII. Finally, section IX concludes our work and gives indications for future enhancements.

II. RELATED AND PREVIOUS WORK

This section gives an overview over existing radiometric compensation techniques. Basic methods are explained first, since they are partially used by our algorithm. Recent approaches that adapt the image content in a pre-process to achieve optimized results with one of the basic methods are described next. In contrast to our algorithm, they can only be applied to static images.

A. Basic Radiometric Compensation Techniques

Most radiometric compensation approaches apply structured light projection and camera feedback for measuring surface and environment parameters, such as geometry, reflectance and environmental light, as well as for establishing a precise mapping between projector and camera pixels. This is usually carried out during a short off-line calibration step and assumes a static constellation afterward. The parameters are used during runtime for radiometric compensation and geometric warping computations. An exception is the work by Fujii et al. [9] that utilizes an co-axial alignment of projector and camera for dynamic compensation on non-static surfaces. A closed feedback loop is used in this case to re-adjust the compensations over time while either surface or projector-camera pair can be moved.

Several algorithms have been developed to project compensated images onto planar and non-planar diffuse surfaces.

Nayar et al. [18] use a 3x3 matrix for each pixel to encode the mixing between the color channels of projector and camera, as well as the surface reflectance. The values of the matrices are measured by projecting a series of uniform color patterns onto the surface and capturing the reflected images. Computing the compensation image is then equivalent to multiplying the inverse color mixing matrices onto the color vectors of the corresponding pixels in the input image. This method (or variations of it) has been applied in combination with planar [18], more complex [12], and non-static [9] surfaces.

Bimber et al. [4] describe a radiometric compensation technique that supports multiple projectors to increase the overall brightness and consequently reduces clipping artifacts on complex surfaces. For planar surfaces - such as paintings - clipping errors can also be minimized by coating the surface with a transparent film material that diffuses a certain amount of light directly while transmitting the remaining portion [2].

Wetzstein et al. [22] present a generalized approach to radiometric compensation by inverting the full light transport captured between projector and camera. This method is extended to support color mixing and the application of an arbitrary number of projectors and cameras. All local and global light modulations, as well as projector and camera defocus can be compensated in real-time with this approach. The calibration process, however, can take up to several hours.

All of these techniques support an image compensation in real-time, but suffer from the same problem: If the compensation image contains values above the maximal brightness or below the black level of the projector, clipping artifacts will occur. These artifacts let the underlying surface structure become visible. While applying multiple projectors [4] is one option that increases the brightness but also the complexity of the projection system, the application of an amplifying film material [2] is another option that restricts such an approach to simple surfaces.

B. Content Dependent Radiometric and Photometric Compensation

Recent algorithms extend radiometric compensation by varying the input image first to achieve an optimized compensation quality with minimized clipping artifacts.

Wang et al. [21] presented the first technique that scales the overall intensity of the input image until clipping errors that result from radiometric compensation are below a perceivable threshold. Their computational intensive numerical minimization can only be applied to static monochrome images and surfaces.

Park et al. [19] describe a technique for increasing the contrast in the compensation image by applying a histogram equalization to the colored input image. This does not preserve the contrast ratio of the original image. Consequently, the image content is thus modified significantly. The problem of occurring clipping errors is not considered by this method.

A complex framework for computing an optimized photometric compensation is presented by Ashdown et al. [1]. The surface's reflectance is scanned with a color calibrated HDR camera. The captured data and the image content is

then transformed into the device-independent CIE L^*u^*v color space in which color distances are based on the human visual perception. The chrominance values are fitted into the gamut of each projector pixel. Finally a luminance fitting is applied with a relaxation method based on differential equations. The compensation algorithm presented in [18] is then used with the adjusted image rather than with the original image. This method can achieve optimized results even for projections onto surfaces with high varying reflectance properties or high saturation - but for static images only.

All of these techniques lead to reduced clipping artifacts and consequently to an increased visual quality compared to the basic compensation methods that do not pre-adapt the input images. However, due to their computational complexity that can mainly be contributed to numerical minimizations, a real-time compensation cannot be achieved. This constrains them to the presentation of still images. Animated content, such as movies, can only be displayed after a time consuming pre-correction. This, however, is impractical for most applications. It is particularly useless if surface and setup don't remain completely static, such as it is the case for mobile projectors that will require a flexible and frequent re-calibration on different surfaces. Furthermore, distributed content, such as DVDs or broadcasted media cannot be pre-corrected for multiple individual surface. Finally, it is clear that real-time dynamic content, such as interactive applications cannot be presented at all.

III. OVERVIEW

Our algorithm performs content adaptation and radiometric compensation in real-time, and reduces visual artifacts while preserving a maximum brightness and contrast. Although we chose the basic compensation scheme presented in [4], only minor modifications to the adaptation algorithm are necessary to use any of the other techniques instead. The algorithm is implemented entirely on the GPU in five steps:

- 1) An analysis of the input image and of the projection surface is performed to gain sufficient parameters for adapting the input image.
- 2) The intensities of the input image are globally scaled depending on the parameters determined in step 1.
- 3) The scaled image that results from step 2 is analyzed for clipping errors.
- 4) The intensities of the image content is re-scaled globally and locally depending on the results of the error analysis in step 3.
- 5) The re-scaled image from step 4 is radiometrically compensated and projected.

Additional time-dependent adaptation factors are applied to avoid abrupt changes in the displayed brightness. These steps will be described in detail below.

IV. ANALYSIS

Generating radiometrically compensated images without clipping errors while simultaneously preserving a high contrast ratio and brightness requires the analysis of both - projection surface and input image. While the captured surface

reflectance has to be analyzed only once¹, the image content has to be analyzed continuously and in real-time. Our algorithm stores each input image in the texture memory of the graphics card. All following tasks are accomplished via fragment shading and multi-pass rendering entirely on the GPU. Framebuffer Objects (FBOs) are used for an efficient exchange of data.

A. Surface Analysis

By applying the basic radiometric compensation technique described in [4], structured light projection and camera feedback delivers several surfaces properties after a one-time off-line calibration step. First, the projector-camera pixel correspondences are determined and stored in two-dimensional look-up tables that are passed as textures to fragment shaders. During run-time, the shaders carry out a per-pixel displacement mapping of all projector pixels to warp the images' geometry with respect to the surface geometry. The result is an undeformed appearance from the perspective of the calibration camera.

For radiometric calibration the same method requires two additional parameters for each pixel:

- The contribution of the (uncontrollable) environmental light which is reflected from the surface - EM (including the projector's black-level).
- The surface's reflectance and the projector-to-surface form-factors - FM (the fall-off of projected intensity, depending on the projector-to-surface distance and the projection angle).

The intensity range for which a radiometric compensation without clipping is possible can now be computed from the two parameters FM and EM . Figure 1 visualizes the

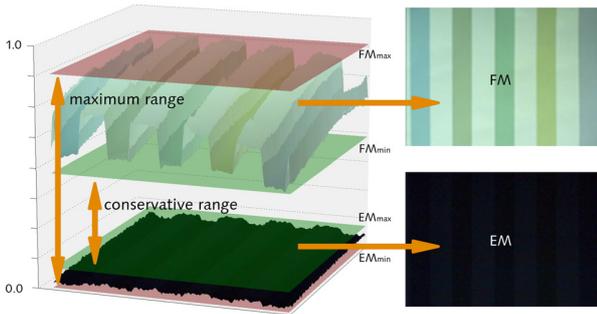


Fig. 1. Three-dimensional view of the intensity range reflected by a striped wall paper. The area between both green planes depict the range of a conservative compensation without clipping errors. The area between the red planes represents the maximum range in which a compensation is possible (potentially with clipping errors).

reflection properties for a sample surface. By analyzing the responses in both datasets, we can compute the range of intensities for a conservative compensation. Thus, any input pixels $O(x, y)$ within this global range (bound by the two

¹As long as we assume that the surface and the projector-camera system remains static.

green planes - from the maximum value EM_{max} to the minimum value FM_{min}) can be compensated correctly for each point on the surface without causing clipping artifacts:

$$EM_{max} \leq O(x, y) \leq FM_{min} \quad (1)$$

The red planes in figure 1 define the maximum range of displayable intensities in which a compensation without clipping is only possible if the color values of any input pixel lie within their local range:

$$EM(x, y) \leq O(x, y) \leq FM(x, y) \quad (2)$$

The calculation of the extreme values is carried out once on the CPU. To avoid extreme values, intensities outside the threefold of the standard deviation are omitted.

B. Content Analysis

Since pixels outside the displayable range cause clipping artifacts, the input image is analyzed to support subsequent global and local luminance adjustments that ensure an optimized compensation. These image processing steps have to be performed in real-time for each input image and are implemented on the GPU to benefit from the high memory bandwidth and parallel processing capabilities of modern graphics hardware.

1) *Average Image Luminance*: The arithmetic luminance mean L_{avg} of the input image can be used as an initial factor for an automatic scaling of the image content. This is described in section V-A. It is computed via multi-pass rendering on the GPU:

A fragment shader transforms the input image into its CIEXYZ representation according to the sRGB transformation matrix. The Y values storing the required luminance information [8] is extracted from this image and is directly rendered into a texture by using FBOs. Throughout multiple rendering passes, the luminance image is successively downscaled by the factor 2 while averaging the four neighboring pixels in each step. The results are also directly rendered into textures via FBOs and are forwarded to the next rendering pass. Like for mip-mapping, this is repeated until the remaining luminance image contains a single pixel that stores the average value. With a slight modification, the same technique can be applied to determine global maximum and minimum values of an image, as required for the error analysis described in section V-B. Using multiple render targets, all of these steps can be carried out in parallel.

2) *Threshold Map*: The perception of luminance variations in images depends on many different factors, such as the display brightness, the local image contrast and the spatial frequencies of the content.

In [20] the so-called threshold map is introduced that contains the maximum non-perceivable luminance threshold for every pixel of an image. This value contains the maximum luminance differences that can be varied in the original image without causing a visually perceivable difference. For computing the threshold map, the physical luminance values

of the displayed image, as well as the spatial frequencies and contrast relations of the image content are required. The luminance-dependent values are calculated separately by using the threshold-versus-intensity (TVI) function and are stored in the so-called TVI map. This function describes the peak luminance sensitivity of the human visual system which is only correct for environments with uniform background luminance. This, however, is not the case if an image with non-uniform content is projected. In this situation the sensitivity depends also on the local image content. These spatially varying factors are computed and stored in the so-called elevation factor map. The threshold map is computed by multiplying the corresponding entries of the TVI map and the elevation factor map.

Thus, varying the local luminance of the input image by an amount that is below the corresponding values in the threshold map will not lead to perceivable luminance differences. A compensated projection, however, can be enhanced if this is done for regions in which clipping occurs. This is described in section V-C.

The TVI map, the elevation factor map and the threshold map are computed on the GPU in real-time, as outlined in the following sub-sections.

a) TVI Map: As proposed in [20], we use the procedure presented in [14] to calculate the TVI map which stores the results of the TVI-function for each pixel. In this approximation, the adaptation luminance L_{adapt} at each pixel is used which is calculated by averaging the luminance values over 1° of the visual angle centered at the according pixel. As described in section IV-B.1, the RGB values of the input image are transformed into their luminance representation first. We use a photometer to measure the average minimum and maximum reflected physical luminance values of the surface by projecting a complete white and black image. These values are then used to transform the image's luminance representation into their physical luminance values. Furthermore, the size of the projection area and the distance to the observer has to be known.

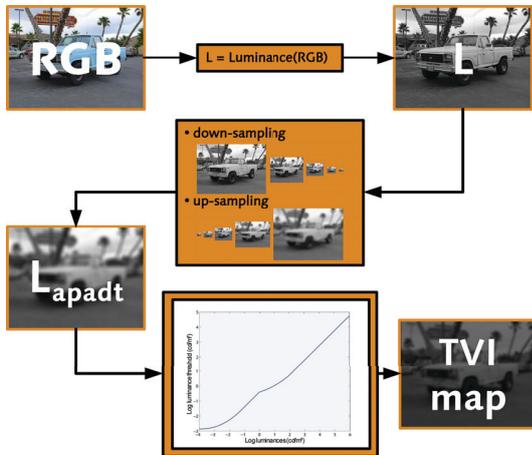


Fig. 2. Flow chart of the GPU-based TVI map generation.

Due to the fact that the TVI function requires the adaptation

luminance which can be reached by smoothing the luminance values over 1° visual angle, the calculated luminance L has to be convolved with a low-pass filter. This smoothing operation can be approximated efficiently via multi-pass rendering on the GPU: First, the luminance image is bilinearly filtered and down-sampled iteratively until a single pixel represents the size of 1° of the visual angle. The result is then iteratively up-sampled and bilinearly filtered until the original resolution is reached. Once again, FBOs are used to render directly into floating point textures. Finally the TVI function is calculated for each pixel, as described in [20], and stored in an additional floating point texture.

Figure 2 illustrates this process. The function in the lower box represents the values of the used TVI-function. Note, that orange boxes represent single or groups of fragment shaders.

b) Elevation Factor Map: The elevation factor map is used for adjusting the TVI map depending on the spatial frequencies and contrast ratios of the input image content. Therefore, a Laplacian pyramid is computed from the input images' Gaussian pyramid. The Laplacian pyramid is then converted into a contrast pyramid that can be used for calculating the elevation factor map as described in [20].

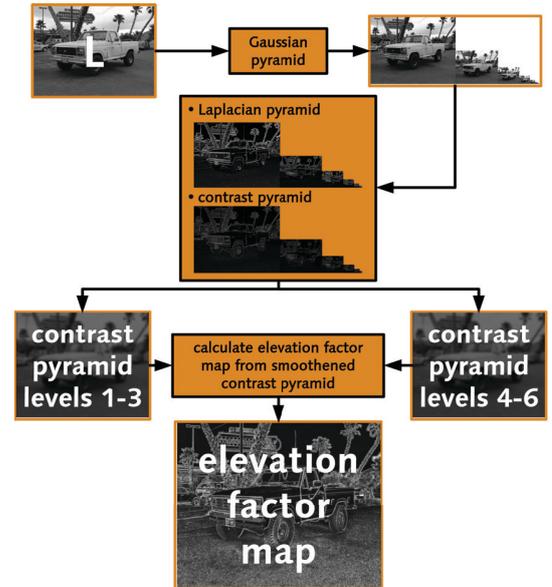


Fig. 3. Flow chart of the GPU-based elevation factor map generation.

Figure 3 outlines the single steps of the GPU-based generation of the elevation factor map.

The Gaussian pyramid is computed from the luminance image by iteratively down sampling and applying a Gaussian blur fragment shader. Another shader is then used for computing the levels of the Laplacian and contrast pyramids.

Since the levels of the contrast pyramid have to be locally averaged next (as suggested by Lubin [16]), they are efficiently stored in the RGB channels of two floating point textures. These computations can be carried out in parallel by using multiple texture targets. The final shader applies the average

operation to each level of the contrast pyramid and calculates the elevation factor map values as discussed in [20].

Corresponding values of the TVI map and the elevation factor map are finally multiplied with a fragment shader to compute the threshold map. Once more, FBOs are used for all of these operations to render directly into floating point textures.

3) *High Spatial Frequencies*: The amount of high spatial frequencies can be approximated by analyzing the Laplacian pyramid. This is a useful parameter for the local intensity variation, described in section V-C. Therefore, the lowest level of the Laplacian pyramid is binarized with a fragment shader based on a pre-defined threshold. Using hardware accelerated occlusion queries the pixels in the high frequency domain are counted. Dividing this by the number of all pixels in the image leads to the ratio of high-frequency to low-frequency pixels.

V. ADAPTATION AND COMPENSATION

With the results of the analysis, the input image is adapted to minimize clipping errors while preserving a high overall brightness and contrast after radiometric compensation. The adaptation is realized in three steps:

- 1) Global scaling of the image's intensities.
- 2) Error analysis of the scaled images resulting from step 1.
- 3) Global and local intensity adjustments based on the errors determined in step 2.

After these adaptation steps, the radiometric compensation is applied and the result is projected. These steps are described in more detail below.

A. Pre-Adaptation

In a first adaptation step the acquired information about the average image luminance and the surface properties (see sections IV-B.1 and IV-A) is used to apply an approximate global scaling of the image's intensity. A compensation image is then calculated from the scaled input image. This allows to analyze the resulting quality of the global intensity adjustments and to identify local clipping errors. The results are used for calculating the final global and local scaling parameters.

The pre-adaptation step is carried out in a reduced resolution to speed up the necessary calculations. The lower sampling rate leads to minimal clipping artefacts that can be tolerated².

The intensities of the input image are scaled depending on the average image luminance and the maximum and minimum color values of the projection surface. While images with a low average luminance are up-scaled, too bright images are down-scaled in their intensity according to equations 3 and 4:

$$\begin{aligned} I_{max}(x, y) &= scale(I(x, y), FM_{max}, EM_{min}) \\ I_{min}(x, y) &= scale(I(x, y), FM_{min}, EM_{max}) \\ scale(in, max, min) &= min + (in \cdot (max - min)) \end{aligned} \quad (3)$$

²We used a resolution of 128² pixels which seems to be a good trade-off between performance and quality on current GPUs. By varying this resolution, the algorithm's performance can be adjusted to the capabilities of the applied graphics hardware.

$I_{min}(x, y)$ and $I_{max}(x, y)$ are two scaled representations of the input image $I(x, y)$. While $I_{min}(x, y)$ represents the image in the conservative compensation range, $I_{max}(x, y)$ is scaled to the maximum range in which a compensation is possible (see section IV-A). The final image intensities $I_{scale}(x, y)$ are a linear interpolation between of $I_{minR}(x, y)$ and $I_{max}(x, y)$. Thereby, the interpolation weights depend on the derived luminance average of the input image L_{avg} :

$$\begin{aligned} I_{scale}(x, y) &= interp(I_{max}(x, y), I_{min}(x, y), L_{avg}) \\ interp(x, y, u) &= x \cdot (1.0 - u) + y \cdot u \end{aligned} \quad (4)$$

After the global adjustment, a radiometric compensation according to [4] is applied to I_{scale} ³. The result is an initial compensation image I_c . A fragment shader analyzes each pixel of I_c for the maximum clipping error in each color channel. The result of this analysis is written into separate color channels of an FBO. Values that are above the highest displayable intensity (i.e., > 1.0) are written into the red color channel, and values that are below the projector's black level are written into the green color channel. If no clipping appears at a pixel, the blue channel is used to store the minimum distance of the pixel's color values to 1.0. Otherwise zero value are stored in this channel. Consequently, the error definitions for each pixel are given with:

$$\begin{aligned} Err_r &= \begin{cases} 1.0 - \max(I_{c,r}, I_{c,g}, I_{c,b}); (I_{c,r} \vee I_{c,g} \vee I_{c,b} < 0) \\ 0; else \end{cases} \\ Err_g &= \begin{cases} |\min(I_{c,r}, I_{c,g}, I_{c,b})|; (I_{c,r} \vee I_{c,g} \vee I_{c,b} < 0) \\ 0; else \end{cases} \\ Err_b &= \begin{cases} 1.0 - \max(I_{c,r}, I_{c,g}, I_{c,b}); (I_{c,r} \vee I_{c,g} \vee I_{c,b} < 0) \\ 0; else \end{cases} \end{aligned} \quad (5)$$

As mentioned above, this step is realized by a direct render-to-texture operation via frame buffer objects. In the next step, the generated error texture Err is analyzed to re-scale the image locally and globally for achieving optimized compensation results.

B. Error Analysis

Clipping errors lead to abrupt alternations in luminance and chrominance within the displayed image.

A conservative global luminance reduction leads to a full elimination of clipping errors, but also to a significant reduction in contrast and brightness. Therefore, our algorithm varies the image intensities locally in addition to a neutralization of remaining clipping errors while preserving a high overall image brightness and contrast.

Studies of human visual perception indicate that abrupt changes in luminance are perceived more intense than smooth and low frequent modifications [17] [10] [15]. Consequently, we blur the calculated clipping errors with a Gaussian smoothing kernel G :

$$\begin{aligned} Err_{FM}(x, y) &= Err_r(x, y) \otimes G(\sigma) \\ Err_{EM}(x, y) &= Err_g(x, y) \otimes G(\sigma) \end{aligned} \quad (6)$$

Attention has to be paid when applying the smoothing operator to the clipping errors: On the one hand, a smooth

³See equations 11 and 12.

local modification is required for avoiding abrupt intensity variations. On the other hand the image content should not be alternated more than necessary. We adjust the σ parameter of the filter kernel inverse proportionally to the amount of high spatial frequencies of the input image (see section IV-B.3). Thus, local luminance reductions affect a larger region in the image if the image content stores mostly low spatial frequencies. The affected area is decreased for a larger amount of high spatial frequencies. For this purpose a series of GPU-based Gaussian filters with varying σ values (ranging from 6 to 16) are applied. The factor that is used for global intensity scaling⁴ equals the average image luminance L_{avg} . It can now be adjusted more precisely with respect to the largest detected clipping value within the image:

$$S' = \min(1, L_{avg} + \max(\max(Err_r) - Err_{max})) \quad (7)$$

If Err_r stores values above a maximum tolerated clipping error Err_{max} , the new global scaling factor S' is increased. This leads to a larger reduction in brightness according to equation 4. Empirical experiments confirmed that an Err_{max} of 0.55 delivers adequate results, while S' is constrained to an upper limit of 1.0 since no clipping can occur in this range ($I(x, y)$ is scaled to the conservative range).

If there is no clipping at all within the entire image, S' will be adjusted with respect to the smallest value in Err_b , which leads to an increase in brightness:

$$S' = \max(0, L_{avg} - 1 + \max(Err_b)) \quad (8)$$

To avoid a perceivable flickering of the projection due to its continuous adjustment, the scaling factor is smoothed over time (see section V-D).

Figure 4 summarizes all calculation steps described above: From the result of the pre-adaptation (a) the computed clipping values are stored in the color channels of an auxiliary texture (b). The red (b_1) and green (b_2) channels store the clipping errors for values above and below the displayable range. Similar to the calculation of the average image luminance described in section IV-B.1, the maximum values in both channels are determined. Depending on these extrema, the global scaling factor is adjusted. Both error maps (b_1 and b_2) are blurred depending on the amount of high spatial frequencies within the input images (c_1 and c_2).

C. Final Adaptation and Compensation

In the remaining adaptation step, the image content is globally re-adjusted as well as locally adapted before computing the compensation image which is finally being projected. For the global re-adjustment, equation 4 is applied together with the adjusted scaling factor S' . The result is the image I'_{scale} . Finally, I'_{scale} is adapted locally, depending on the blurred clipping errors. Therefore, Err_{FM} and Err_{EM} have to be up-scaled to projector resolution. To avoid an unnecessary re-scaling in areas in which no clipping occurs and to keep the perceivable image manipulation at a minimum, the luminance values $L(x, y)$ and the threshold map $TM(x, y)$ are also

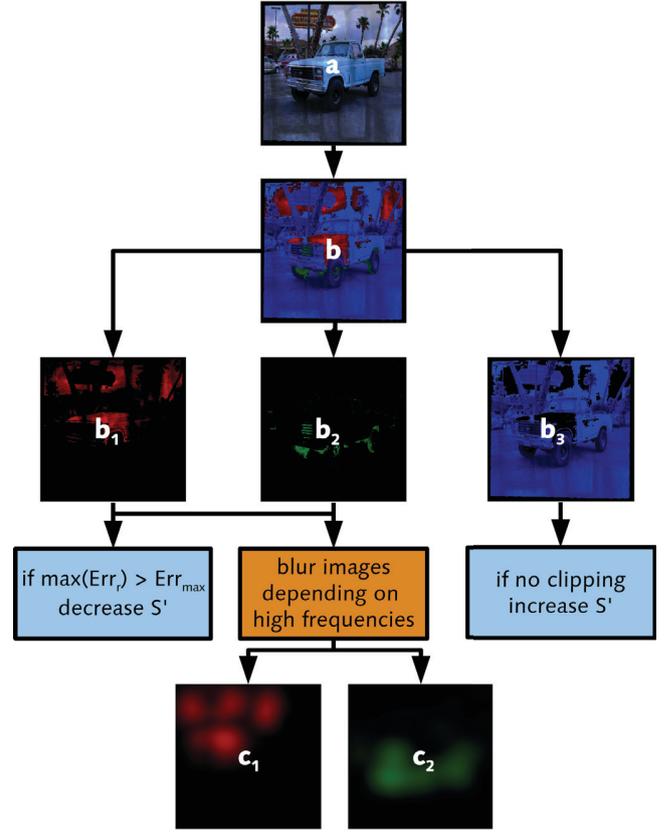


Fig. 4. Flow chart describing the steps of the clipping error analysis.

considered. The local adaptation in areas with clipping values above 1.0 is achieved with:

$$l = L(x, y) \cdot Err_{FM}(x, y) \cdot f_1$$

$$I''_{scale}(x, y) = \begin{cases} I'_{scale}(x, y) - l, & l < TM(x, y) \cdot f_2 \\ I'_{scale}(x, y) - TM(x, y) \cdot f_2, & else \end{cases} \quad (9)$$

In this case local intensities are decreased in the globally scaled image content. Clipping errors that occur in the reverse case are due to values below the black-level of the projector. Consequently, the local image intensities have to be increased:

$$l = (1.0 - L(x, y)) \cdot Err_{EM}(x, y) \cdot f_1$$

$$I''_{scale}(x, y) = \begin{cases} I'_{scale}(x, y) + l, & l < TM(x, y) \cdot f_2 \\ I'_{scale}(x, y) + TM(x, y) \cdot f_2, & else \end{cases} \quad (10)$$

$I''_{scale}(x, y)$ stores the results of the final adaptation stage. The scaling factor f_1 can be used for varying the local adaptation manually. An optimal value of $f_1 = 1.5$ was found empirically. It represent a good trade-off between quality and performance.

Blurring the error maps causes a spatial distribution of clipping error values. This may lead to unnecessary large intensity modifications in the original image. To adapt the amount of modification to the actual brightness of a pixel, the clipping errors Err_{FM} and Err_{EM} are weighted by the corresponding luminance values $L(x, y)$.

The threshold map can also be adjusted manually with a second scaling factor f_2 . We found that $f_2 = 2.0$ was a good compromise between a perceivable local reduction and

⁴Equation 4.

an efficient minimization of visible clipping. If clipping entries are stored in $Err_{FM}(x, y)$ and in $Err_{EM}(x, y)$, the multiplied luminance values $L(x, y)$ guarantee that both adjustments do not cancel each other out. This is because values in $Err_{FM}(x, y)$ affect only input pixels with high intensities, while values in $Err_{EM}(x, y)$ have only an impact on pixels with low intensities.

The adapted input image can now be radiometrically compensated in such a way that the resulting image is displayed with maximal brightness and contrast while disturbing clipping artifacts are minimized. As mentioned earlier we use the compensation equation presented in [4] for a one projector setup:

$$I'_c = \frac{I''_{scale} - EM}{FM} \quad (11)$$

The same adaptation algorithm can be applied, if n overlaying projectors ($p = 1..n$) are used to produce a brighter image at the surface. In this case the surface analysis has to be carried out under simultaneous illumination of all projectors. The radiometric compensation equation then extends to (details can be found in [4]):

$$I'_{c,p} = \frac{I''_{scale} - \sum_{p=1}^n E_p M}{\sum_{p=1}^n F_p M} \quad (12)$$

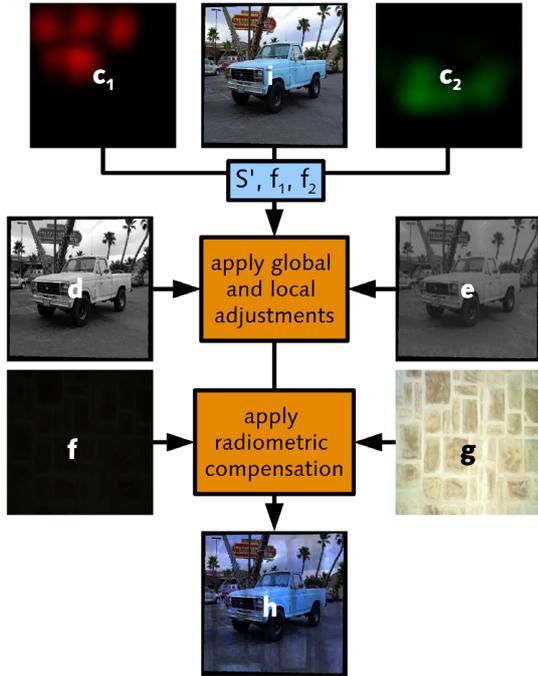


Fig. 5. Flow chart of the final image adaptation and radiometric compensation steps.

Figure 5 summarizes the steps of the final image adaptation and the subsequent radiometric compensation.

The adjusted scaling factor S' that results from the error analysis is used for re-scaling the input image globally. The smoothed error textures (c_1 and c_2) are applied together with the threshold map (e) and the local luminance information (d) to perform local intensity adjustments in the input image (i).

The threshold map constraints the local intensity variations with respect to the largest non-perceivable luminance. In the last step, the surface reflectance (g) and the environment light contribution (f) are required to compute the adapted compensation image (h) that is projected onto the surface.

D. Temporal Adaptation

Our algorithm adjusts each input image individually. In case of animated content, such as videos or interactive applications, this might lead to abrupt changes in brightness and contrast and to visible flickering.

To avoid these disturbing effect, we adapt the global scaling factors over time - depending on factors used for scaling previous images. We apply a temporal adaptation model that was initially developed for interactive tone mapping [7] [11] [13]:

$$S''_i = S'_{i-1} + (S'_i - S'_{i-1}) \cdot (1 - e^{-\frac{T}{\tau}}) \quad (13)$$

The temporally adapted scaling factor S''_i depends on the factor S'_i computed for the current frame, and on the factor S'_{i-1} used for the previous frame. In addition, an exponential attenuation function that is determined based on the actual frame rate T , as well as a constant τ that describes the rate of human luminance adaptation are taken into account. Projection displays usually operate with luminance values in the range of photopic vision. Thus, a value of $\tau = 0.1$ for rods is used [13]. With this temporal adaptation, abrupt global luminance variations are converted to smooth intensity changes over time. In some cases clipping might happen shortly. However, it can hardly be perceived since the adjustments are carried out very fast. The blurred clipping errors (Err_{FM} and Err_{EM}) are smoothed over time to decrease the visibility of the local intensity adjustments.

Instead of using the computed error textures Err_i at time instance i , the values are averaged with the error values used for the previous image Err'_{i-1} :

$$Err'_i(x, y) = \frac{Err'_{i-1}(x, y) + Err_i(x, y)}{2} \quad (14)$$

This is computationally less expensive than applying equation 13 to each individual error pixel.

VI. EXAMPLES

In this section, we want to provide several visual examples of our algorithm's outcome⁵.

Figure 6 illustrates a projection (a) onto a striped wall paper (a). The projection surfaces contains gray scales in this first example. An uncompensated projection (b) leads to a clear visibility of the underlying surface. In (c) an intermediate result of our adaptive algorithm without local reductions is shown. While it appears to be very similar to the original image (e), clipping artifacts are visible (especially in the bright area of the upper left corner). The final result is shown in (d). It includes automatic global and local adaptations. Visible clipping errors are reduced without decreasing the image's overall contrast and brightness much.

⁵See video for dynamic examples.



Fig. 6. Adaptive radiometric compensated projection with global (c) and additional local luminance adjustments (d).



Fig. 7. Projection onto a striped wallpaper with saturated colors, without (c) and with (d) adaptive radiometric compensation.

Figure 7 demonstrates a similar example as in figure 6. This time, however, an image (a) is projected onto a surface with saturated colors (b). Although slight color mismatches compared to the original image are detectable, our algorithm produces an acceptable result (d) - especially when comparing it with an uncompensated projection (c).

Figure 8 demonstrates another example. This time an image (e) is projected onto a wooden surface (a). With the local intensity adjustments, our algorithm produces a brighter result (d) compared to a manually adjusted compensation (c).

Note, that in all examples presented in this section, the original images are always illustrated in their native digital format, while the compensation results are shown as photographs of the projections onto the individual surfaces. Consequently, differences between original images and projected compensations can be contributed to camera parameters, such as response,



Fig. 8. Projection onto a wooden surface, without (b), with manually adjusted (c) and with adaptive radiometric compensation (d).

resolution, angle, distance, and field of view.

The main advantage of our approach over basic radiometric compensation algorithms can be shown with animated or interactive content. The continuous adaptation of the input images leads to a constant improvement of image quality compared to a compensations with manual adjustments that do not adapt to the displayed content.

Figure 9 illustrates two different frames from the movie Shrek 2⁶, projected onto a natural stone wall (a). While (b) contains bright scenes, a dark scene is shown in (e). The two frames are approximately one second apart in the original video. As demonstrated in (c,f), a basic compensation algorithm (e.g., [4]) will fail in this situation. On the one hand, visible clipping errors occur in image areas with bright intensities (c) due to the physical limitations of the projector. On the other hand, the displayed image becomes too dark (g) due to static adjustment parameters of the basic method. Similar results will be received with all other un-adaptive radiometric compensation methods. As illustrated in (d,g), our adaptive approach responds to these situations automatically and reduces the visibility of the underlying surface while widely preserving brightness and contrast of the original video.

VII. PERFORMANCE ANALYSIS

Related algorithms [21] [1] implement an adapted radiometric compensation by applying numerical minimizations and relaxation algorithms. In contrast to our method, they are not capable of achieving interactive or real-time frame rates. This, however, is essential for future applications of radiometric compensation in general. This sections provides an analysis of our algorithm's performance.

On our test platform⁷ a PAL-resolution video can be compensated with approximately 35 frames per second.

⁶©DreamWorks Animation SKG.

⁷Intel Pentium 4, 2.8 GHz, 1 GB RAM, NVidia GeForce 7900 GTX, XGA projector resolution.

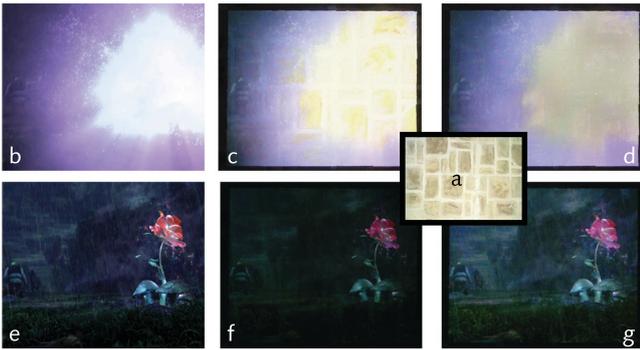


Fig. 9. Two frames of an animation (b, e) projected onto a natural stone wall with a static radiometric compensation (c,f) and with our adaptive algorithm (d,g).

Since the resolution that is chosen for the pre-adaptation step has a significant impact on the overall performance, it can be adjusted according to the desired frame-rate. Figure 10 illustrates a diagram of the measured performance that can be achieved for different pre-adaptation resolutions. As mentioned earlier, we chose a pre-adaptation resolution of 128×128 pixels, since it proved to be a good trade-off between performance and image quality for our hardware. As explained in section V-C the results of the pre-adaptation step are linearly interpolated and up-scaled to the resolution of the projector before the final adaptation step is carried out.

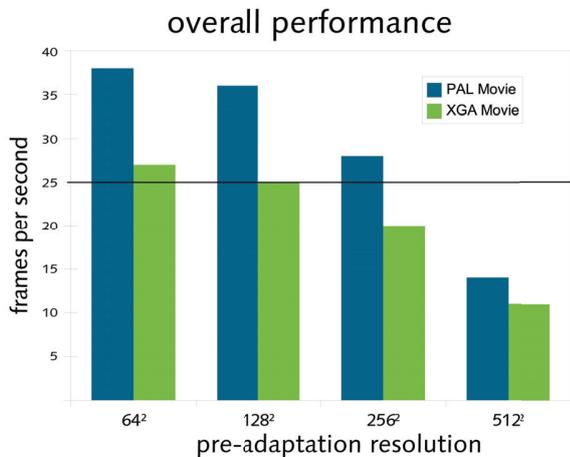


Fig. 10. Measured performance for different pre-adaptation resolutions.

Table I outlines the advantage of a GPU implementation compared to an optimized CPU implementation. It compares four of the necessary image processing operations⁸ in our shader implementations with corresponding CPU realizations of Intel's OpenCV image processing library.

In all four tasks the shader computations on the GPU outnumber the CPU implementations. On our test platform we archived speed-up factors of 1.7 to 38.7. Furthermore, additional memory transfers between GPU and CPU are necessary for a CPU implementation. They have not been considered

⁸Gaussian pyramid, average calculation, Gaussian blur and high frequency calculation.

TABLE I
COMPARISON BETWEEN CPU AND GPU IMPLEMENTATIONS OF FOUR NECESSARY IMAGE PROCESSING TASKS (IN FPS).

Task	CPU	GPU	Gain factor
Gaussian pyramid	38	166	4.3
Gaussian blur (Kernel: 77^2 pixels)	4	115	38.7
Average luminance	175	303	1.7
High frequency analysis	41	397	9.6

at this point. We believe that GPUs will become a regular component of modern video projectors in the near future.

VIII. INFORMAL USER STUDY

An informal user study was carried out to validate the increase of perceived visual improvement that can be gained by an adaptive approach compared to a basic radiometric compensation. Therefore participants were asked to compare and to judge the visual appearance of projected sample sequences. A preference for our adaptive radiometric compensation over basic methods was expected.

The user study was performed in a dark room without environment light⁹. Each subject had to adapt to the lighting conditions for five minutes before the test sequences were presented to them.

Two LCD-projectors¹⁰ were used for the user study. They were color calibrated with a photometer for fitting the gamuts of both devices. While the first projector was used to project radiometrically compensated images onto a natural stone wall, the second one projected the original image onto a white canvas.

The user study was separated into three parts: In the first part a series of still images was projected sequentially onto the stone wall with a duration of 15 seconds. Each image was shown next to each other at the same time¹¹ - one compensated with the adaptive algorithm, the other one compensated with a constant basic method [4]. For the latter case, the same intensity scaling factor was used for all images. It was chosen in such a way that it equals the average scaling factor generated by the adaptive algorithm for all presented test images. Whether the adaptive algorithm was used on the left or the right side was selected at random to avoid an influence of the slightly different underlying surface. The images were labeled to avoid confusion. The subjects did not know which image was generated by which algorithm. They were asked to compare both and decide which one is preferred over the other one.

In the second part of the user study, four video sequences were projected one after another. Again, both compensations (adaptive and static) were presented next to each other, and their positions were randomly switched. For this experiment, two different scaling factors were used for the static compensation method: The first factor was chosen to avoid visible clipping errors completely. This leads to dim, but clipping-free

⁹Except the black level of the projector.

¹⁰A Sony VPL-CX80.

¹¹One on the left side, the other one on the right side.

projections. Videos number 1 and number 2 were presented this way. The second factor was selected to be equal the average scaling factor of the adaptive algorithm over all presented video frames. This was used to display videos number 3 and number 4.

In the third part of the user study, the adaptive and static compensations of the still images presented in the first part were compared with a conventional projection onto a white planar canvas. For this task the participants had to decide which of the compensated images appears more like the reference projection.

For all three parts the subjects had to rate their preferences within five scales, ranging from "left image much more convenient" over "no difference" to "right image much more convenient".

For evaluation purposes, these scales were converted to numerical values ranging from -2 to 2 . While positive numbers represent a preference for the adaptive algorithm, negative values indicate a preference for a constant radiometric compensation. Altogether 32 subjects¹² participated in the user study.

While the subjects indicated only a small preference for the adaptive algorithm when still images were presented, it was significantly favored for dynamic content. Especially videos with varying contrast and brightness levels were perceived as enhanced. The diagram in figure 11 illustrates this: A significant preference of the adaptive method was indicated for all four sample videos. It was confirmed that the adaptive approach delivers results that appear more like an ordinary projection than the static compensation method.

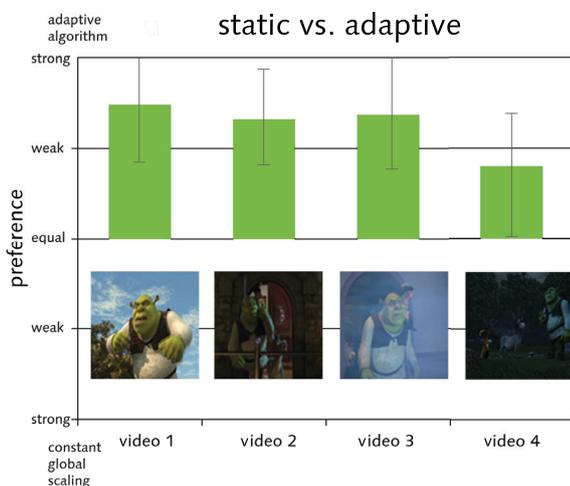


Fig. 11. Comparison of four compensated (static and adapted) videos projected onto a natural stone wall were compared.

IX. CONCLUSION AND FUTURE WORK

In this article we presented a real-time capable adaptive radiometric compensation algorithm to enable enhanced projections onto everyday surfaces. It considers the surface's reflectance and geometry, the image content and the capabilities

of the human visual system to reduce visible artifacts that are due to the limited dynamic range and brightness of projectors. We believe that such techniques will be essential to support the portability of pocket projectors and projection systems that are integrated into mobile devices, such as cell-phones, PDAs and digital cameras.

Our algorithm analyses the projection surface as well as the image content and adapts the input images globally as well as locally in two steps before a radiometric compensation is applied. This leads to a minimization of clipping errors and corresponding chrominance shifts while preserving a maximum of brightness and contrast. The adaptation parameters are temporally adjusted to ensure smooth intensity transitions and to avoid visible flickering. While a GPU implementation of our algorithm enables real-time frame rates, a user study has confirmed that an automatic adaptation leads to higher visual image quality compared to basic compensation methods.

Besides a luminance adaptation, chrominance adjustments are future extensions to our algorithm that enable brighter projections on highly saturated surfaces. Currently all clipping errors are blurred with the same Gaussian filter kernel. Image segmentation will allow to compute and apply individual kernel parameters for different clipping areas. Both of these enhancements will lead to an improved image quality - but on the cost of performance. An acceptable balance has to be found. As in many other visualization areas, a corpus of performance bottlenecks will be solved by upcoming graphics hardware.

REFERENCES

- [1] M. Ashdown, T. Okabe, I. Sato and Y. Sato, "Robust Content-Dependent Photometric Projector Compensation," *Proc. IEEE International Workshop on Projector-Camera Systems*, pp. 60-68, 2006
- [2] O. Bimber, F. Coriand, A. Kleppe, E. Bruns, S. Zollmann and T. Langlotz, "Superimposing Pictorial Artwork with Projected Imagery," *IEEE MultiMedia volume 12*, pp. 16-26, 2005
- [3] O. Bimber, G. Wetzstein, A. Emmerling and C. Nitschke, "Enabling View-Dependent Stereoscopic Projection in Real Environments," *Proc. of International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 14-23, 2005
- [4] O. Bimber, A. Emmerling and T. Klemmer, "Embedded entertainment with smart projectors," *IEEE Computer 38*, pp. 56-63, 2005
- [5] N. Boukala, P. Colatini and J.D. Runga, "Fast and accurate color image processing using 3d graphics cards," *Proc. of 8th International Fall Workshop: Vision Modeling and Visualization*, pp. ,2003
- [6] R. Dumont, F. Pellacini and J.A. Ferwerda, "A Perceptually-Based Texture Caching Algorithm for Hardware-Based Rendering," *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, pp. 249-256, 2001
- [7] F. Durand and J. Dorsey, "Interactive Tone Mapping," *Proceedings of the Eurographics Workshop on Rendering Techniques*, pp. 219-230 ,2000
- [8] J.D. Foley, A. v.Dam., S.K. Feiner, J.F. Hughes, *Computer Graphics: Principles and Practice*, Addison Wesley, Reading, MA, second edition, 1990
- [9] K. Fujii, M.D. Grossberg, S.K. Nayar, "A Projector-Camera System with Real-Time Photometric Adaptation for Dynamic Environments," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'05, Volume 01*, pp. 814-821, 2005
- [10] E.B. Goldstein, *Sensation and Perception*, Wadsworth Publishing Company, 2001
- [11] N. Goodnight, R. Wang, C. Woolley and G. Humphreys, "Interactive time-dependent tone mapping using programmable graphics hardware," *Proceedings of the 14th Eurographics workshop on Rendering*, pp. 26-37, 2003
- [12] M.D. Grossberg, P. Harish and S.K. Nayar, "Making One Object Look Like Another: Controlling Appearance Using a Projector-Camera System," *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 452-459, 2004

¹²Mixed female and male, between 21 and 36 years of age.

- [13] K. Grzegorz, K. Myszkowski and H.-P. Seidel, "Perceptual Effects in Real-Time Tone Mapping," *Proceedings of the 21st spring conference on Computer graphics*, pp. 195-202, 2005
- [14] G. W. Larson, H. Rushmeier and C. Piatko, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 3, No. 4, pp. 291-306, 1997
- [15] C. J. Lloyd, "Quantifying edge blend quality: Correlation with observed judgements," *Proceedings of Image Conference*, 2002
- [16] J. Lubin, "A Visual Discrimination Model for Imaging System Design and Evaluation," *Vision Models for Target Detection and Recognition*, World Scientific, pp. 245-283, 1995
- [17] A. Majumder, "A Practical Framework To Achieve Perceptually Seamless Multi-Projector Displays," *Ph. D. Thesis, Department of Computer Science, UNC-Chapel Hill*, 2003
- [18] S. K. Nayar, H. Peri, M. D. Grossberg and P. N. Belhumeur, "A Projection System with Radiometric Compensation for Screen Imperfections," *Proc. IEEE International Workshop on Projector-Camera Systems*, pp. , 2003
- [19] H. Park, M.-H. Lee, S.-J. Kim and J.-Il Park, "Contrast Enhancement in Direct-Projected Augmented Reality," *ICME'06: International conference on Multimedia and Expo*, 2006
- [20] M. Ramasubramanian, S. Pattanaik and D.P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 7382, 1999)
- [21] D. Wang, I Sato, O. Takahiro and Y. Sato, "Radiometric Compensation in a Projector-Camera System Based Properties of Human Vision System," *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, Volume 03, page 100*, 2005
- [22] G. Wetzstein and O. Bimber, "Radiometric Compensation of Global Illumination Effects with Projector-Camera Systems," *SIGGRAPH '06: Proceedings of the 33th annual conference on Computer graphics and interactive techniques, Poster*, 2006

A Generalized Approach to Radiometric Compensation

GORDON WETZSTEIN and OLIVER BIMBER
Bauhaus-University Weimar

Video projectors have evolved tremendously in the last decade. Reduced costs and increasing capabilities have led to widespread applications in home entertainment and visualization. The rapid development is continuing. Projector-camera systems enable a completely automatic calibration. Novel image compensation techniques that allow seamless projections onto complex everyday surfaces have recently been proposed. They support the presentation of visual content in situations where projection-optimized screens are not available or not desired - as in museums, historic sites, air-plane cabins, or stage performances. Furthermore, the anticipated mobility enabled through laptops and pocket projectors will also imply the possibility to use on-site surfaces for presentations instead of carrying projection screens. So far, existing compensation methods consider only local illumination effects, such as diffuse reflections - which is sufficient for many situations. Global illumination effects, such as inter-reflections, refractions, scattering, etc. are ignored. We propose a novel method that applies the light transport matrix for performing an image-based radiometric compensation which accounts for all possible types of light modulation. For practical application the matrix is decomposed into clusters of mutually influencing projector and camera pixels. The compensation is modeled as a linear system that can be solved with respect to the projector patterns. Precomputing the inverse light transport in combination with an efficient implementation on the GPU makes interactive compensation rates possible. Our generalized method unifies existing approaches that address individual problems. Based on examples, we show that it is possible to project corrected images onto complex surfaces such as an inter-reflecting statuette, glossy wallpaper, or through highly-refractive glass. Furthermore, we illustrate that a side-effect of our approach is an increase in the overall sharpness of defocused projections.

Categories and Subject Descriptors: I.3.3 [Computer Graphics]: Picture/Image Generation; I.4.9 [Image Processing and Computer Vision]: Applications

General Terms: Theory

Additional Key Words and Phrases: Projector-Camera Systems, Radiometric Compensation, Inverse Light Transport

1. INTRODUCTION

Rapid advances in electro-mechanics and optics have increased the capabilities of projectors in terms of spatial resolution, brightness, dynamic range, throw-ratio, and speed. Cost reductions, availability, and the fact that projectors (in contrast

This work was accomplished at the Augmented Reality Laboratory of the Bauhaus-University Weimar. Some of the results were initially presented as a research poster at SIGGRAPH 2006. Authors' addresses: G. Wetzstein and O. Bimber, Media Faculty, Bauhaus-University Weimar, Bauhausstrasse 11, 99423 Weimar, Germany; email: {wetzstein,bimber}@medien.uni-weimar.de. Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0730-0301/20YY/0100-0001 \$5.00

to flat panels) can display images that are much larger than the devices themselves made them a mass-market product. Besides for (home-)entertainment, education and business purposes, high resolution tiled-screens and immersive surround screen projections are used for visualizations of scientific, engineering and other content. Emissive CRT technology is losing more and more ground to light-valve technology, like LCD, LCOS and especially DLP.

The trend toward a higher flexibility of multi-projector systems with respect to device configuration and screen alignment is well noticeable. Recently, numerous projector-camera approaches that enable a seamless projection onto complex everyday surfaces have been proposed. In general this is referred to as *radiometric compensation*. These techniques correct the projected images for geometrical distortions, color and intensity blending, and defocus caused by the underlying surface. Eventually, the final images appear as being projected onto a planar white canvas - even though this is not the case.

Previously proposed radiometric compensation techniques assume a simple geometric relation between cameras and projectors that can be automatically derived using structured light projections or co-axial projector-camera alignments. This results in a precise mapping between camera and projector pixels. In reality, the light of a projected pixel often bounces back and forth several times at different areas on the surface, before it eventually reaches the imaging sensor of the camera. Due to inter-reflection, refraction, scattering and other global illumination effects, multiple camera pixels at spatially distant regions on the image plane may be affected by a single projector pixel. A direct mapping usually considers only camera pixels with the highest intensity contribution that result from the captured light of corresponding modulated projector pixels. Consequently, all global illumination effects are discarded. In some cases it might not even be possible to acquire a direct mapping at all because global effects are too dominant.

We propose a novel image-based approach to radiometric compensation that accounts for all possible local and global illumination effects. Conventional light transport acquisition schemes are employed to capture these effects with projector-camera systems. The goal of our compensation is to find illumination patterns that, when projected onto a complex surface, result in a desired image from the camera's perspective. We model the compensation as a linear system that can be solved with respect to the projected pattern. Due to the size of the resulting equation systems it is necessary to decompose the light transport into clusters of mutually influencing camera and projector pixels. An interactive compensation by solving the system for each frame is in most cases not possible. This, however, can be achieved by pre-computing the inverse light transport, which makes an efficient implementation on the GPU possible. Depending on the complexity of the scene and the occurring global effects, real-time frame rates can be achieved.

2. RELATED WORK

Seamless Multi-Projections: The research on projector-camera systems has recently gained a lot of interest in the computer graphics and the computer vision community. Traditionally, multi-projector configurations are employed to create large-scale high-resolution displays on planar diffuse screens. To achieve a consistent geometrical alignment, as well as a correct photometric and radiometric appearance can be a challenging problem for such systems. Geometric registration can be obtained using homography matrices for planar screens, via projector calibration and projective texture-mapping for non-trivial screens with known geometry, or through look-up tables and per-pixel displacement mapping for complex surfaces with unknown geometry. Photometric correction involves intensity linearization and fitting as well as color gamut matching and cross fading. A good overview of camera-based projector calibration techniques can be found in [Yang et al. 2005].

For projection screens with spatially varying reflectance, radiometric compensation techniques as presented in [Nayar et al. 2003; Grossberg et al. 2004; Bimber et al. 2005; Fujii et al. 2005] can be applied to minimize the artifacts induced by the light modulation between projection and surface pigments. Content-dependent, adaptive radiometric compensation techniques that are optimized to human perception have been described in [Ashdown et al. 2006; Wang et al. 2005; Grundhoefer and Bimber 2006]. All of these methods presume a well-defined mapping between projector and camera pixels.

Focus Related Projector-Camera Techniques: Yet another interesting aspect of projection systems is image focus and defocus. [Bimber and Emmerling 2006] projected images with a large depth of field that are composed from different contributions of multiple overlapping projectors with varying focal planes. The overall sharpness of an image displayed by a single projector was enhanced in [Zhang and Nayar 2006] and [Brown et al. 2006]. Therefore, the defocus kernels of light samples projected onto complex scenes was analyzed. Based on these results, image sharpening was employed to compensate for optical defocus digitally.

Forward Light Transport, BRDF Acquisition and Relighting: The forward light transport between a light source and an imaging device implicitly takes all global illumination effects into account. Recently, it has been used for BRDF and BSSRDF acquisition [Goesele et al. 2004; Peers et al. 2006], image-based relighting [Debevec et al. 2000; Sen et al. 2005; Masselus et al. 2003; Garg et al. 2006] and environment matting [Zongker et al. 1999].

Inverse Illumination: The compensation of scattering for immersive and semi-immersive projection displays with known screen geometry using a reverse radiosity technique was presented in [Bimber et al. 2006]. While the required form factors were precomputed, [Seitz et al. 2005] proposed a technique that estimates global illumination parameters with a camera and a laser pointer for canceling inter-reflections in photographs. The operator that is applied to an image for removing indirect il-

lumination is represented as a matrix. This matrix is a composition of the scene’s inverse light transport and a forward light transport matrix containing only the indirect illumination contributions. Latter theoretically exists for arbitrary scenes, but has only been shown to be measurable for lambertian surfaces.

3. RADIOMETRIC COMPENSATION AS INVERSE LIGHT TRANSPORT

The idealized forward light transport between a projector and a camera is given by

$$\vec{c}_\lambda = T_\lambda \vec{p}_\lambda + \vec{e}_\lambda. \quad (1)$$

This is a well-known equation, where \vec{c}_λ is a single color channel λ of a camera image with resolution mn . It is represented as a column vector of size $mn \times 1$. T_λ is the light transport matrix, which can be acquired through structured illumination as described in [Sen et al. 2005]. \vec{p}_λ is the projector pattern of resolution pq represented as a column vector of size $pq \times 1$, and \vec{e}_λ is the environment light including the projector’s black level captured from the camera (also represented as a $mn \times 1$ column vector). Solving equation 1 for \vec{p}_λ is equivalent to radiometric compensation of all global and local light modulations contained in T_λ .

Due to the spectral transmission properties of the color filters that are used in cameras and projectors, individual spectral components between both devices cannot be fully separated. Projecting red light only, for instance, leads to non-zero responses in the camera’s green and blue color channel. This is known as *color mixing*. Equation 1 is based on the assumption that color mixing between camera and projector is negligible. For deriving a generalized mathematical framework of radiometric compensation it has to be taken into account.

A generalized light transport equation can be formulated that includes color mixing:

$$\begin{bmatrix} \vec{c}_R - \vec{e}_R \\ \vec{c}_G - \vec{e}_G \\ \vec{c}_B - \vec{e}_B \end{bmatrix} = \begin{bmatrix} T_R^R & T_R^G & T_R^B \\ T_G^R & T_G^G & T_G^B \\ T_B^R & T_B^G & T_B^B \end{bmatrix} \begin{bmatrix} \vec{p}_R \\ \vec{p}_G \\ \vec{p}_B \end{bmatrix} \quad (2)$$

Thus, solving equation 2 for $[\vec{p}_R \vec{p}_G \vec{p}_B]^T$ represents the general radiometric compensation of all light modulations for a single-camera single-projector configuration. It is formulated as a linear equation system of the size $3mn \times 3pq$. The subscripts indicate individual color channels. Each single light transport matrix $T_{\lambda_2}^{\lambda_1}$ (size: $mn \times pq$) has three channels for RGB colors, representing the contribution of the projected light to all camera channels λ_2 . The superscripts indicate the light transport matrix that is determined for a specific projector channel λ_1 . Thus T_R^G , for example, is the red color channel of the light transport matrix acquired for the green projector channel.

When describing a direct relation between a single projector and camera pixel, each $T_{\lambda_2}^{\lambda_1}$ is a scalar. In this case, the coefficient matrix is of size 3×3 . This equals the local compensation model used by [Nayar et al. 2003], where the coefficient matrix is referred to as *color mixing matrix* V . However, equation 2 is its generalized form that also takes global light effects into account.

An extension toward general configurations containing an arbitrary number of cameras and projectors can easily be derived from equation 2. Indexing the light transport matrices is done by denoting two additional sub- and superscripts on the left side of T . While the left superscript indicates a projector, the left subscript indicates a camera. Thus, matrix ${}_c^p T_{\lambda_2}^{\lambda_1}$ represents the light transport from projector p to camera c . The radiometric compensation equation for k projectors and r cameras is given by

$$\begin{bmatrix} {}_0\vec{c}_R - {}_0\vec{e}_R \\ {}_0\vec{c}_G - {}_0\vec{e}_G \\ \vdots \\ {}_{(r-1)}\vec{c}_B - {}_{(r-1)}\vec{e}_B \end{bmatrix} = \begin{bmatrix} {}_0T_R^R & {}_0T_R^G & \dots & {}^{(k-1)}T_R^B \\ {}_0T_G^R & {}_0T_G^G & \dots & {}^{(k-1)}T_G^B \\ \vdots & \vdots & \ddots & \vdots \\ {}_0T_B^R & {}_0T_B^G & \dots & {}^{(k-1)}T_B^B \end{bmatrix} \begin{bmatrix} {}_0\vec{p}_R \\ {}_0\vec{p}_G \\ \vdots \\ {}^{(k-1)}\vec{p}_B \end{bmatrix}. \quad (3)$$

It has to be solved for $[{}_0\vec{p}_R \ {}_0\vec{p}_G \ \dots \ {}^{(k-1)}\vec{p}_B]^T$. Employing multiple cameras and projectors allows to capture samples of the full 8D reflectance field. While a compensation for multiple cameras can support view-dependent approaches as in [Bimber et al. 2005], a compensation for multiple projectors can remove shadow casts as well as minimize specular highlights (see [Park et al. 2005]), and allows increasing the overall brightness, dynamic range, resolution and focus of projected images, as known from existing multi-projector techniques. However, all possible local and global illumination effects are considered, and existing techniques are unified in our case.

4. CLUSTERING AND HARDWARE ACCELERATION

Due to its enormous size it is impractical to solve the equation system for the entire projector pattern in one step. A possibility of simplifying this computation is to decompose the light transport matrices into clusters of mutually influencing camera and projector pixels. Each of these clusters represents a single, smaller equation system that can be processed independently.

Global and local light modulation will influence the connections within the matrices differently. A flat diffuse surface with an overall focused projector and camera normally produces many small and localized clusters. Capturing a scene with a large depth variance and global light effects will lead to fewer widely connected clusters.

Neighboring projector pixels are likely to overlap in the camera image. This does not necessarily result from global illumination effects, but can also be contributed to camera or projector defocus, differences in resolution, lens imperfections or sensor specific effects such as blooming. Searching for inter-connections in the scanned light transport often leads to a single or to a few large clusters. Since these clusters form equation systems that are too large to be solved efficiently, connections within the matrix must be removed to decompose it into independent sub-clusters. Therefore, the matrix is represented as a weighted bipartite graph that contains camera pixels and projector pixels as nodes. Dependent projector-camera pixel pairs are connected by edges which are weighted with the corresponding luminance transfer contribution. Figure 1 (c) illustrates a sample graph, where all nodes are directly or indirectly connected. Hence, this graph equals a single connected cluster.

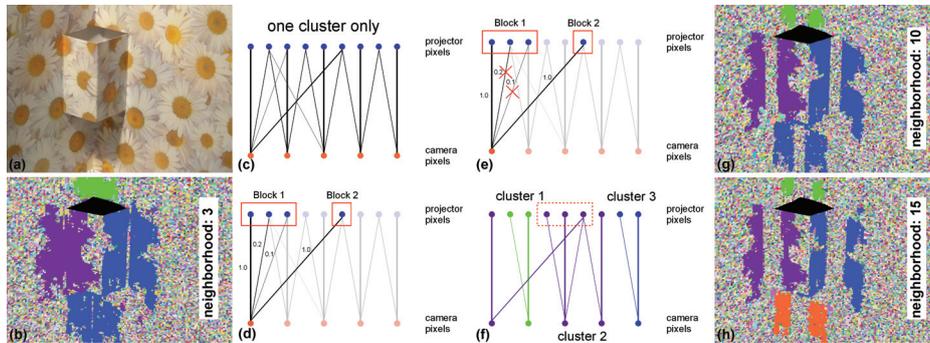


Fig. 1. A camera pixel may be affected by several projector pixels either due to local overlaps or indirect illumination (d). These projector pixels can be grouped into blocks of spatially neighboring pixels in projector space. In order to separate global from local illumination, the contribution for each camera pixel can be restricted per block. A larger neighborhood size leads to smaller cluster sizes (a+b+g+h).

The goal is to split local overlaps while preserving contributions from spatially more distant areas in the projector image which result from global illumination. This can be achieved by grouping all connected projector pixels of a single camera pixel into blocks of spatially neighboring regions in projector space (fig. 1 (d)).

For each camera pixel, blocks are formed by searching the connected projector pixel with the highest edge weight in the graph. This pixel represents the center of a new block. All neighboring pixels in the projector image that are also connected to the same camera pixel are inserted into this block. A constant neighborhood size is used to define the adjacencies. Elements that are assigned to a specific block are not further considered for building new blocks. This is iteratively continued until all connected projector pixels are part of a neighborhood block.

Lower luminance contributions (i.e., edges with lower weights) within each block are then cut out of the graph depending on a given threshold, as it can be seen in figure 1 (e). Searching for new clusters after removing connections leads to a larger

number of smaller clusters (fig. 1 (f)). The process is repeated recursively until each cluster has a predefined size that allows to solve its equations system.

Increasing the neighborhood size is likely to produce smaller and more frequent clusters at the expense of discarding the interaction between close projector pixels as depicted in figures 1 (a+b+g+h)¹.

For radiometric compensation, the camera pixels of each cluster are replaced by pixels of the input image. While \vec{c} and T are known, \vec{p} has to be computed and finally be projected onto the scene. The environment light \vec{e} is a camera image captured under black projector illumination. To avoid negative values in the compensation image \vec{p} , the cluster-individual equation systems can be solved separately using iterative non-negative least squares (NNLS) methods. Figure 2 shows an example².

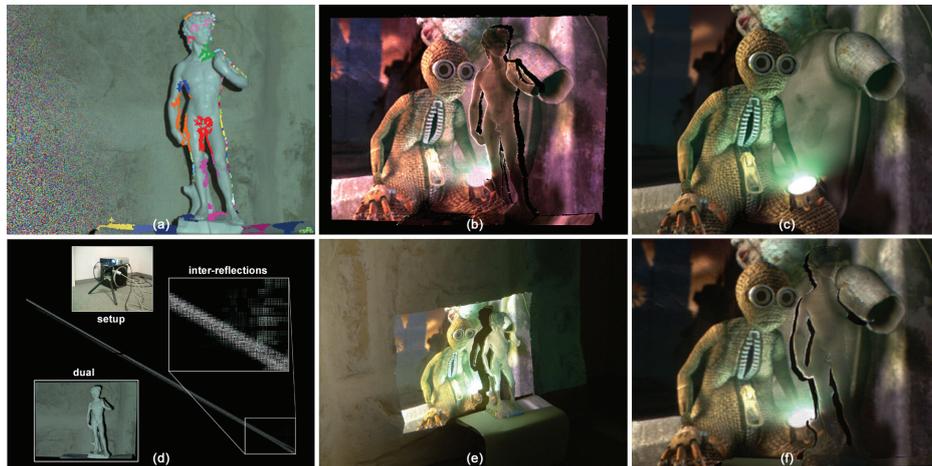


Fig. 2. A floodlight image of a scene containing inter-reflections between the statue's arms and legs partially superimposed with decomposed clusters (a). Applying the light transport matrix visualized in (d) enables to synthesize a compensation image (b) for an input image (c). Projecting this onto the objects (e) results in a corrected view from the camera's perspective (f). Since only one projector was used, shadow casts cannot be filled in the camera view.

Computing a single compensation image by solving multiple equation systems takes several seconds to minutes - depending on the size of the clusters and available computing resources. This makes a real-time compensation of dynamic content, such as movies or real-time graphics impossible.

¹The luminance filter-threshold remained constant for the visualized decompositions.

²The compensation was performed using the idealized light transport (eqn. 1) and took approximately 3 minutes on a P4, 3GHz, 2 GB RAM. The displayed content is a screen shot from the short film "9", Focus Features and 9, LLC.

However, equation 1 can be reformulated by applying the inverse of the light transport matrix to both sides, yielding $T_\lambda^+ (\vec{c}_\lambda - \vec{e}_\lambda) = \vec{p}_\lambda$. Similarly, equations 2 and 3 can be converted. A matrix inverse exists only for regular square matrices. Since this cannot be guaranteed for the light transport matrix of a general projector-camera system its pseudo-inverse has to be determined. This is calculated using SVD, which minimizes the 2-norm of difference between the input image and the result of the radiometric compensation. Each of the pseudo-inverse's columns represents the projector pattern that would have to be projected onto the scene for illuminating only a single camera pixel at a time.

Computing a pseudo-inverse matrix is numerically expensive compared to solving the corresponding equation system explicitly. However, for our case this allows the problem to be split into a computational expensive preprocessing step (computing T^+ for each cluster) and a simple vector dot product of each T^+ 's rows and the input image. The latter step is carried out for multiple projector pixels in parallel on the GPU during runtime. A pseudo-inverse is computed for each cluster and inserted into the appropriate locations of the matrix. All of its non-zero elements are packed into floating point textures for direct access on the GPU. A look-up table provides information about the appropriate matrix elements for each projector pixel.

Comparing the NNLS solutions to the result of the multiplication with T^+ did not reveal visible differences. Slight intensity variations in both solutions are due to numerical instabilities. Computing the pseudo-inverses for the example shown in figure 2 took approximately 15 minutes (P4, 3 GHz, 2 GB RAM), while the compensation was performed with 7 fps on a GeForce 7900 GTX, 512 MB.

5. COMPENSATING LOCAL AND GLOBAL ILLUMINATION EFFECTS

Refraction and other complex light modulations represent a challenging problem for structured light scanning techniques. It is often not possible to determine a precise mapping of individual projector pixels and corresponding camera pixels. Figure 3 (a), for instance, shows a glass in front of a wallpapered surface. The projection of text³ through the glass reveals image distortions that can be attributed to refraction. Multiple characters are visible at different locations within the camera image. These distortions increase near the glass' bottom area.

Geometric image distortions, intensity variations as shadows and caustics, as well as color artifacts from refraction and blending with the background make a radiometric compensation difficult in this example. The off-diagonal branches in the acquired light transport matrix (fig. 3 (b)) clearly indicate the existence of global illumination effects. While the twisted narrow bands (upper right close-up) are due to refractions, the blank portions on the matrix's diagonal (left magnification) represent the thicker parts of the glass' rim that do not reflect light toward the camera. An interesting effect is highlighted in the magnified lower right part of figure 3 (b).

³The poem "Jabberwocky" by Lewis Carroll from the book "Through the Looking-Glass" (1872).

These matrix entries belong to the image of the glass' base that is visible to the camera only because of reflections at the bottom of the glass. Note that this area cannot be fully compensated because it also reflects other parts of the scene that are not illuminated by the projector. In general, portions of the camera image that are not lit by the projector (such as the darker parts of the glass' shadows on the background) cannot be compensated. Employing multiple projectors can account for such shadow regions.

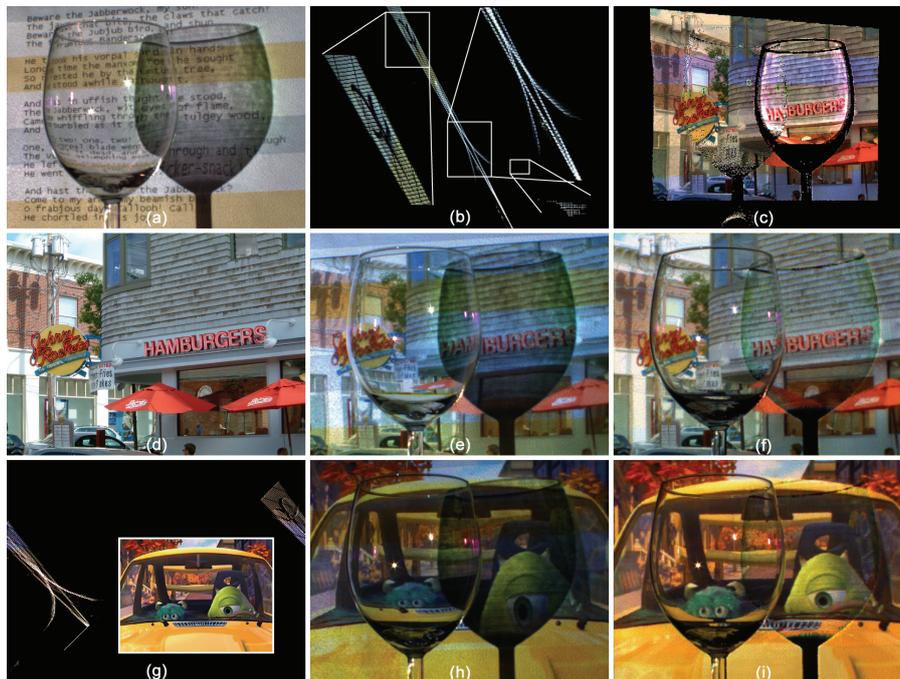


Fig. 3. A glass in front of a colored wallpaper (a). The acquired light transport matrix (b) is decomposed and used to compute a compensation image (c) for an input image (d). The result captured from the camera's point of view for the uncompensated (e) and the compensated (f) projection. Applying the inverse light transport matrix (g) allows a real-time compensation for displaying interactive content and movies - uncompensated (h) and compensated (i).

Figure 3 illustrates real-time⁴ compensation examples for a static photograph and a movie sequence (g)⁵. It also demonstrates several physical limitations of radiometric compensation techniques in general. If the same surface is visible multiple times in the camera image (e.g., due to a refracted/reflected and a direct view) it is

⁴The matrix's pseudo-inverse was computed in app. 13 minutes (P4, 3 GHz, 2GB RAM). About 30 fps were achieved for compensation with the GPU implementation (GeForce 7900 GTX, 512 MB).

⁵From the short film "Mike's New Car", courtesy Pixar).

generally not possible to compute a single compensation value for strongly different input responses. These cases are implicitly optimized by solving the equation systems in a least squared error sense (e.g., via with NNLS).

The second example, presented in figure 4, shows a scene containing two V-shaped cardboard pieces in front of a glossy wallpaper (fig. 4 (a)). The left one is coated with a self-adhesive transparent film. Diffuse scattering and inter-reflections lead to an increased brightness and to color bleeding in the corner areas (fig. 4 (b+e)). Performing a radiometric compensation to compensate these effects, as shown in figures 4 (c+f), the differences⁶ are depicted in figures (d+g). Figures 4 (h) and (j) show an uncompensated and a compensated projection of a movie frame (i)⁷.

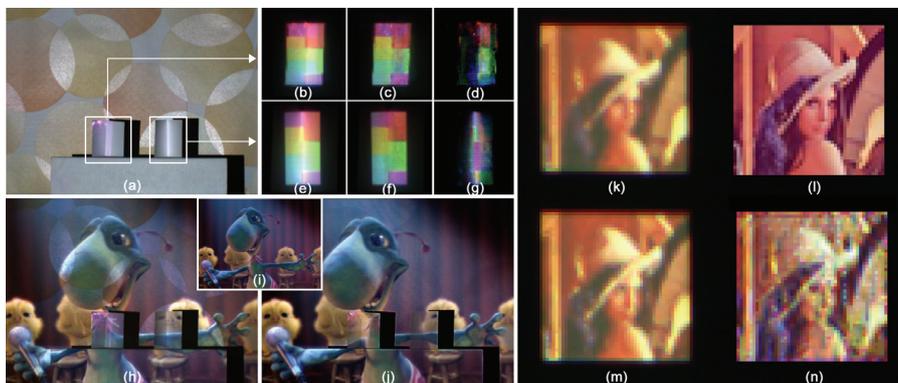


Fig. 4. The inverse light transport enables to compensate diffuse scattering (e+f+g), glossy reflections and inter-reflections (b+c+d), and to increase the perceived focus (n). Uncompensated projections of the original images (i+l) are shown in (h+k), while (m) is a manually sharpened projection of (l).

Because of their large optical apertures and resultant narrow depth of field, conventional video projectors can display focused images on single fronto-parallel planes only. Projecting onto surfaces with large depth variance leads to regionally defocused images. Since the projector defocus is also included in the light transport matrix, blurred projections are implicitly sharpened by the radiometric compensation. This is demonstrated in figures 4 (k-n)⁸. An uncompensated projection of a picture (l) with a defocused projector on a planar uniformly colored surface is shown in (k). The desired image is very small (48x48 pixels), however, the resulting equation system contains 2304x2700 elements. About 25 minutes were required for solving this with a non-negative least squared error solution (P4, 3GHz, 4 GB

⁶The images are contrast enhanced by 75%.

⁷From the short film "The Chubb Chubbs", courtesy Pixar.

⁸Note that the light transport matrix is not decomposed for this experiment, however, only affected projector and camera pixels are included in the equation system.

RAM) on the CPU.

Previously described techniques such as [Zhang and Nayar 2006] and [Brown et al. 2006] employed the measured defocus kernels of the specific setup to estimate an optimal sharpening of the projection image. Due to the optical defocus of the projector, the captured image appeared more similar to the original image than its unmodified projection. The idea of focus optimization through pre-sharpening is illustrated in figure 4 (m) while figure 4 (n) is the result of our radiometric compensation. Although only a standard sharpening operator (as found in most image processing applications) was used, the captured projection appears more focused. The limits of sharpening are set by the actual defocus of the projector and of the original image. Thus, it is not possible to compensate optical defocus if the original image does not contain a minimal amount of digital blur.

6. DISCUSSION AND OUTLOOK

Acquiring the forward light transport between a projector and a camera allows capturing the entire light modulation of a projected pattern within an arbitrarily complex scene - including all local and global illumination effects. It represents a 4D slice of the 8D reflectance field and can be described as a linear equation system. Replacing the camera image in the forward light transport equation with a desired picture enables to perform radiometric compensation by solving this equation system for the corresponding projector pattern.

The single-projector single-camera forward light transport equation can be extended to support multi-projector multi-camera setups, as well as color mixing between the devices. This leads to a large equation system that addresses the general case of the full 8D reflectance field. As in the 4D case, this equation system has to be solved for the projected radiance based on the expected irradiance on the camera sensor for radiometric compensation.

To achieve this in practice, the transport matrix is decomposed into a set of independent clusters that form smaller equation systems which can finally be processed. Solving the set of equation systems explicitly on the CPU, however, does not support a compensation in real-time. Pre-computing the inverse light transport matrix on the CPU and performing a simple matrix-vector multiplication during run-time on the GPU does lead to interactive frame-rates.

With this general approach it becomes possible to compensate a variety of local and global illumination effects with a single technique. This unifies a pallet of existing methods that all address individual problems with specialized techniques (e.g., [Nayar et al. 2003; Zhang and Nayar 2006; Brown et al. 2006; Bimber and Emmerling 2006; Bimber et al. 2005; Bimber et al. 2006; Park et al. 2005; Seitz et al. 2005]). Examples for inter-reflections, refractions, diffuse scattering, and the sharpening of defocused projections have been provided in this article. However, it can be expected that other modulation types, such as diffractions, sub-surface

scattering, or specular highlights will follow the same scheme. All experiments have been carried out with single-camera single-projector setups. In order to validate our theory, multi-projector multi-camera configurations have to be implemented and evaluated in the future.

As described in [Bimber et al. 2005], multiple sample cameras can be used along with image-based rendering and interpolation techniques to support a view-dependent radiometric compensation of local illumination effects for moving observers. Combining this with the generalized approach that is described in this article eventually results in a light field like rendering technique that performs radiometric compensation. If the light transport is known for multiple sample cameras, rays in between can be synthesized and rendered in real-time (see [Levoy and Hanrahan 1996]). In order to compensate view-dependent local and global illumination effects such as specular reflections or refractions, many sample cameras are necessary. However, the light transport from a single projector to multiple cameras can be acquired simultaneously. Hence, the overall acquisition time does not increase much.

The discussed generalized approach reveals several general limitations of radiometric compensation. Shadows and view-dependent effects such as specular reflections may not be compensatable with a single-camera single-projector configuration. Employing multiple projectors and cameras, however, can account for these situations. The overall focal depth, brightness and resolution are increased as well. However, it is impossible to find an exact solution when a single projector pixel is mapped to multiple camera pixels with different values in the input image (i.e., if the same surface portion is visible multiple times in the camera image). This can be the result of reflections, refractions, or other global effects.

Applying the inverse light transport, either by solving the equation system with NNLS or by precomputing the transport matrix's pseudo-inverse with SVD and multiplying it with the input image, minimizes the squared error between input image and provided solution in the first, case and the 2-norm of difference between both images in the second case.

Finally, the limited brightness, resolution, contrast and in particular the relatively high black level contribution of conventional LCD or DLP projectors represent technical limitations that currently prevent from making all surfaces types disappear completely when applying radiometric compensation.

REFERENCES

- ASHDOWN, M., OKABE, T., SATO, I., AND SATO, Y. 2006. Robust Content-Dependent Photometric Projector Compensation. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS 2006)*.
- BIMBER, O. AND EMMERLING, A. 2006. Multi-Focal Projection: A Multi-Projector Technique for Increasing Focal Depth. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*. Vol. 12.
- BIMBER, O., EMMERLING, A., AND KLEMMER, T. 2005. Embedded Entertainment with Smart Projectors. In *IEEE Computer*. 56–63.
- ACM Transactions on Graphics, Vol. V, No. N, Month 20YY.

- BIMBER, O., GRUNDHOEFER, A., ZEIDLER, T., DANCH, D., AND KAPAKOS, P. 2006. Compensating Indirect Scattering for Immersive and Semi-Immersive Projection Displays. In *IEEE Virtual Reality (IEEE VR'06)*.
- BIMBER, O., WETZSTEIN, G., EMMERLING, A., AND NITSCHKE, C. 2005. Enabling View-Dependent Stereoscopic Projection in Real Environments. In *International Symposium on Mixed and Augmented Reality (ISMAR'05)*. 14–23.
- BROWN, M., SONG, P., AND CHAM, T.-J. 2006. Image Pre-Conditioning for Out-of-Focus Projector Blur. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*.
- DEBEVEC, P., HAWKINS, T., TCHOU, C., DUKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the Reflectance Field of a Human Face. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 145–156.
- FUJII, K., GROSSBERG, M., AND NAYAR, S. 2005. A Projector-Camera System with Real-Time Photometric Adaptation for Dynamic Environments. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. 814–821.
- GARG, G., TAVALA, E.-V., LEVOY, M., AND LENSCH, H. P. A. 2006. Symmetric Photography: Exploiting Data-sparseness in Reflectance Fields. *Eurographics Symposium on Rendering (ESGR)*.
- GOESELE, M., LENSCH, H. P. A., LANG, J., FUCHS, C., AND SEIDEL, H.-P. 2004. DISCO: Acquisition of Translucent Objects. *ACM Trans. Graph.* 23, 3, 835–844.
- GROSSBERG, M., PERI, H., NAYAR, S., AND BELHUMEUR, P. 2004. Making One Object Look Like Another: Controlling Appearance using a Projector-Camera System. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. I. 452–459.
- GRUNDHOEFER, A. AND BIMBER, O. 2006. Real-Time Adaptive Radiometric Compensation. In *SIGGRAPH Poster*.
- LEVOY, M. AND HANRAHAN, P. 1996. Light Field Rendering. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM Press, New York, NY, USA, 31–42.
- MASSELUS, V., PEERS, P., DUTRÉ, P., AND WILLEMS, Y. D. 2003. Relighting with 4D Incident Light Fields. *ACM Trans. Graph.* 22, 3, 613–620.
- NAYAR, S. K., PERI, H., GROSSBERG, M. D., AND BELHUMEUR, P. N. 2003. A Projection System with Radiometric Compensation for Screen Imperfections. In *International Workshop on Projector-Camera Systems (PROCAMS 2006)*.
- PARK, H., LEE, M.-H., KIM, S.-J., AND PARK, J.-I. 2005. Specularity-free projection on nonplanar surface. In *Proceedings of Pacific-Rim Conference on Multimedia (PCM)*. 606–616.
- PEERS, P., VOM BERGE, K., MATUSIK, W., RAMAMOORTHI, R., LAWRENCE, J., RUSINKIEWICZ, S., AND DUTRÉ, P. 2006. A Compact Factored Representation of Heterogeneous Subsurface Scattering. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 25, 3 (July).
- SEITZ, S. M., MATSUSHITA, Y., AND KUTULAKOS, K. N. 2005. A Theory of Inverse Light Transport. In *ICCV*. 1440–1447.
- SEN, P., CHEN, B., GARG, G., MARSCHNER, S. R., HOROWITZ, M., LEVOY, M., AND LENSCH, H. P. A. 2005. Dual Photography. *ACM Trans. Graph.* 24, 3, 745–755.
- WANG, D., SATO, I., OKABE, T., AND SATO, Y. 2005. Radiometric Compensation in a Projector-Camera System Based on the Properties of Human Vision System. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS 2005)*.
- YANG, R., MAJUMDER, A., AND BROWN, M. 2005. Camera Based Calibration Techniques for Seamless Multi-Projector Displays. *IEEE Transactions on Visualization and Computer Graphics* 11, 2.
- ZHANG, L. AND NAYAR, S. K. 2006. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*.
- ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment Matting and Compositing. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 205–214.

Passive-Active Geometric Calibration for View-Dependent Projections onto Arbitrary Surfaces

Stefanie Zollmann, Tobias Langlotz, Oliver Bimber

Bauhausstrasse 11

Bauhaus University Weimar

99425 Weimar

Tel.: +49 (0)3643 583724

Fax: +49 (0)3643 583709

E-Mail: {zollmann, langlotz, bimber}@medien.uni-weimar.de

Abstract: In this paper we present a hybrid technique for correcting distortions that appear when projecting images onto geometrically complex, colored and textured surfaces. It analyzes the optical flow that results from perspective distortions during motions of the observer and tries to use this information for computing the correct image warping. If this fails due to an unreliable optical flow, an accurate –but slower and visible– structured light projection is automatically triggered. Together with an appropriate radiometric compensation, view-dependent content can be projected onto arbitrary everyday surfaces. An implementation mainly on the GPU ensures fast frame rates.

Keywords: dynamic geometric calibration, projector-camera systems, optical flow

1 Introduction and Motivation

Projecting images onto surfaces that are not optimized for projections becomes more and more popular. Such approaches will enable the presentation of graphical, image or video content on arbitrary surfaces. Virtual reality visualizations may become possible in everyday environments - without specialized screen material or static screen configurations (cf. figure 1). Upcoming pocket projectors will enable truly mobile presentations on all available surfaces of furniture or papered walls. The playback of multimedia content will be supported on natural stonewalls of historic sites without destroying their ambience through the installations of artificial projection screens.

Several real-time image correction techniques have been developed that carry out geometric warping [BiWe2005], radiometric compensation [BiEm2005][NaPe2003], and multi-focal projection [BiEm2006] for displaying images on complex surfaces without distortions.



Figure 1: View-dependent stereoscopic projection of 3D content onto large natural stonewall.

As long as the geometry of a non-trivial (e.g. multi-planar), non-textured surface is precisely known, the geometric warping of an image can be computed. Projecting the pre-warped image from a known position ensures the perception of an undistorted image from a known perspective (e.g., of a head-tracked observer or a calibrated camera). Hardware accelerated projective-texture mapping is a popular technique that has been applied for this purpose several times [RasBro1999]. As soon as the surface geometry becomes geometrically more complex or the surface is textured, projective texture-mapping will fail due to imprecisions in the calculations. Minimization errors lead to small misregistrations between projector-pixels and corresponding surface pigments. In addition, projective texture-mapping models a simple pinhole geometry and does not consider the lens distortion of the projector optics. All of this leads to calibration errors in the order of several pixels, and finally to well visible blending artifacts of compensated pixels that are projected onto wrong surface pigments.

Several techniques have been introduced that ensure a precise pixel-individual warping of the image by measuring the mapping of each projector pixel to the corresponding pixels of a calibration camera when being projected onto a complex surface [BiEm2005][PoAlt1982]. Structured light projection techniques are normally being used for determining these correspondences. This leads to pixel-precise look-up operations instead of imprecise image warping computations. The surface geometry does not have to be known. However, since the look-up tables are only determined for one perspective (i.e., the perspective of the calibration camera), view-dependent applications are usually not supported. But this becomes essential for supporting moving observers.

In this paper we describe an online image-based approach for view-dependent warping of images being projected onto geometrically and radiometrically complex surfaces. It continuously measures the image distortion that arises from the movement of an observer by computing the optical flow between the distorted image and an estimated optimal target image. If the distortion is low, the optical flow alone is used for image correction. If the distortion becomes too high, a fast structure light projection is automatically triggered for recalibration.

2 Related Work

Camera-based geometric calibration techniques can be categorized into online or offline methods. While an offline calibration determines the calibration parameters (such as the projector-camera correspondence) in a separate step before runtime, an online calibration performs this task continuously during runtime. A lot of previous work has been done on offline calibration – but little on online techniques.

Active offline calibration techniques usually rely on structured light projection to support enhanced feature detection [PoAlt1982] [CaKi1998]. For simple surfaces with known geometry the geometric image warping can be computed with beforehand determined constant calibration parameters. Examples are homography matrices (for planar surfaces [YanGo2001] [CheSu2002]) or intrinsic and extrinsic projector parameters for non-trivial, non-complex surfaces with known geometry [RasBro1999]. These techniques also support a moving observer since the image warping is adapted to the users position in real-time. For geometrically complex and textured surfaces with unknown geometry, projector-camera correspondences can be measured offline for a discrete number of camera perspectives. During runtime, the correct image warping is approximated in real-time by interpolating the measured samples depending on the observer’s true perspective [BiWe2005].

Online techniques can apply imperceptible structured light patterns that are seamlessly embedded into the projected image. This can be achieved by synchronizing a camera to a well-selected time-slot during the modulation sequence of a DLP projector [CoNa2004]. Within this time-slot the calibration pattern is displayed and detected by the camera. Since such an approach requires modifying the original colors of the projected image, a loss in contrast can be an undesired side effect. Other techniques rely on a fast projection of images that cannot be perceived by the observer. This makes it possible to embed calibration patterns in one frame and compensate them with the following frame. Capturing altering projections of colored structured light patterns and their complements allows the simultaneous acquisition of the scene’s depth and texture without loss of image quality [WaWü2005].

A passive online method was described for supporting a continuous autocalibration on a non-trivial display surface [YanWe2001]. Instead of benefiting from structured light projection, it directly evaluates the deformation of the image content when projected onto the surface. This approach assumes a calibrated camera-projector system and an initial rough estimate of the projection surface to refine the reconstructed surface geometry iteratively.

3 Our Approach

This section describes our online calibration approach. It represents a hybrid technique, which combines active and passive calibration.

3.1 Initialization

For initializing the system offline a calibration camera must be placed at an arbitrary position – capturing the screen surface. The display area on the surface can then be defined by outlining the

two-dimensional projection of a virtual canvas, as it would be seen from the calibration camera's perspective. The online warping approach ensures that –from a novel perspective– the corrected images appear as to be displayed on this virtual canvas. It also tries to minimize geometric errors that result from the underlying physical (non-planar) surface.

For the initial calibration camera the pixel correspondence between camera pixels and projector pixels is determined by projecting a fast point pattern. This results in a two-dimensional look-up table that maps every projector-pixel to its corresponding camera-pixel. This look-up table can be stored in a texture and passed to a fragment shader for performing a pixel displacement mapping [BiEm2005]. The warped image is projected onto the surface and appears undistorted from the perspective of the calibration camera.

In addition the surface reflectance and the environment light contribution are captured from the calibration camera position. These parameters are also stored in texture maps, which are passed to a fragment shader that carries out a per-pixel radiometric compensation to avoid color blending artifacts when projecting onto a textured surface [BiEm2005].

3.2 Fast Geometric Calibration

For determining the correspondences between projector and camera pixels, a fast point pattern technique is used (cf. figure 2).

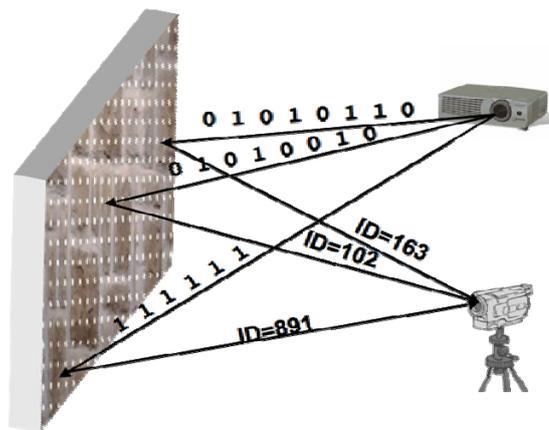


Figure 2: Active registration using binary coded point patterns.

A grid of n points is projected simultaneously. Thereby, each point can be turned on or off – representing a binary 1 or 0 (cf. figure 3-left). Projecting a sequence of point images allows transmitting a binary code at each grid position optically from projector to camera. Each code represents a unique identifier that establishes the correspondence between points on both image planes – the one of the projector and the one of the camera. Depending on the resolution of the grid the code words differ in length. Thus, a minimum of $ld(n)$ images have to be projected to differentiate between the n grid points. Additional bits can be added for error-detection.

To create a continuous lookup table for each projector pixel from the measured mappings of the discrete grid points, tri-linear interpolation is being applied. To benefit from hardware acceleration of programmable graphic cards this lookup table is rendered with a fragment shader into a 16bit texture. The resulting *displacement map* stores the required x,y -displacement of each projector pixel in the r,g channels of the texture. To achieve a further speed-up, spatial coding can be used instead of a simple binary pattern. Projecting two distinguishable patterns per point (e.g., a circle and a ring) allows encoding three states per position and image (cf. figure 3-right). For this case, the minimal number of projected images required to encode n points drops to $\log_2(n)/2$.

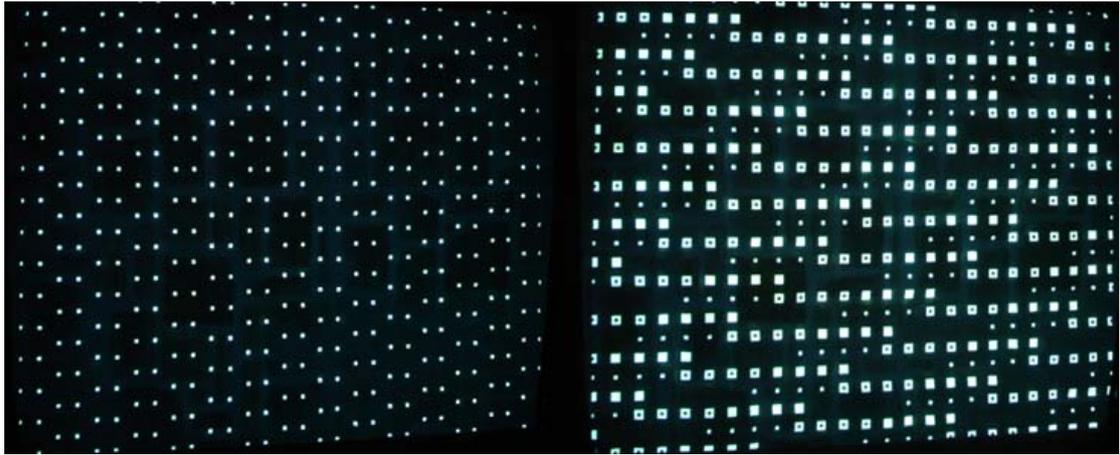


Figure 3: Binary coding (left) and spatial coding (right).

We found that more sophisticated coding schemes (e.g., using color, intensities or more complicated spatial patterns) are difficult to differentiate reliably when projecting onto arbitrarily colored surfaces. This is in particular the case if off-the-shelf hardware is being applied. With a conventional and unsynchronized camcorder (175ms latency) we can scan 900 grid points in approximately one second (by sending two bits of the codeword per projected pattern).

3.3 Passive Calibration for Small Perspective Changes

Once the system is calibrated we can display geometrically corrected and radiometrically compensated images as long as projector and camera/observer are stationary [BiEm2005]. As soon as the observer moves away from the sweet spot of the calibration camera, geometric distortions become perceivable. Note that as long as the surface is diffuse, radiometric distortions do not arise from different positions.

For the following we assume that the camera is attached to the observer's head – matching his/her perspective. This can be realized by mounting a lightweight pen camera to the worn stereo goggles (in case active stereo projection is supported [BiWe2005]). Consequently, the resulting distortion can be continuously captured and evaluated. As described in section 3.1, a pixel correspondence between the initial calibration camera C_0 and projector P exists. For a new

camera position (C_i) that results from movement, another pixel correspondence between C_i and C_0 can be approximated based on optical flow analysis. A mapping from C_i to P is then given over C_0 . Thus, two look-up tables are used: one that stores camera-to-camera correspondences $C_i \rightarrow C_0$, and one that holds the camera-to-projector correspondences $C_0 \rightarrow P$.

Our algorithm can be summarized as follows and will be explained in more detail below:

```
1:     if camera movement occurs
2:         if camera movement stops
3:             calculate optical flow between initial ...
               ...camera image and current camera image
4:             filter optical flow vectors
5:             calculate homography from optical flow
6:             transform default image with homography
7:             calculate optical flow between current...
               ...camera image and transformed image
8:             calculate displacement from optical flow
9:         endif
10:    endif
```

Online calibration algorithm.

For computing $C_i \rightarrow C_0$ the system tries to find feature correspondences between the two camera perspectives (C_i and C_0) and computes optical flow vectors. This step is triggered only when the observer stops moving in a new position (line 2), which is characterized by constant consecutive camera frames.

Since the image captured from C_i is geometrically and perspectively distorted over the surface (cf. figure 4b), the correct correspondences $C_i \rightarrow C_0$ cannot be determined reliably from this image. Therefore, a corrected image for the new perspective C_i has to be computed first. Since neither the camera nor the observer is tracked this image can only be estimated. The goal is to approximate the perspective projection of the virtual canvas showing the geometrically undistorted content from the perspective of C_i (cf. figure 4d). Thereby, the virtual canvas has to appear as to be static in front or behind the surface. Two steps are carried out for computing the perspectively corrected image (line 3-6) – an image plane transformation followed by a homography transformation:

First, the original image content is texture mapped onto a quad that is transformed on the image plane of C_0 in such a way that it appears like an image projected onto a planar surface seen from the initial camera position (line 3). Note, that this first transformation does not contain the any perspective distortion. It is being carried out initially to increase the quality of feature tracking algorithms during the following optical flow analysis. Consequently, the optical flow of a discrete number of detectable feature points between the transformed texture and the image captured from C_i can be computed on the same image plane. Unreliable optical flow vectors are

filtered out (line 4). The remaining ones allow us to compute displacements for a discrete set of feature points between the distorted image from C_i and the undistorted image from C_0 .

The second step uses these two-dimensional correspondences for computing a homography matrix that transforms the geometrically undistorted image from the perspective of C_0 into the perspective of C_i to approximate the missing perspective distortion. Since the homography matrix describes a transformation between two viewing positions ($C_0 \rightarrow C_i$) over a plane, this transformation is only an approximation, which is sufficient to estimate the ideal image at the new camera position. To avoid an accumulation of errors the homography matrix that approximates the mapping between C_0 and the new camera position C_i is applied.

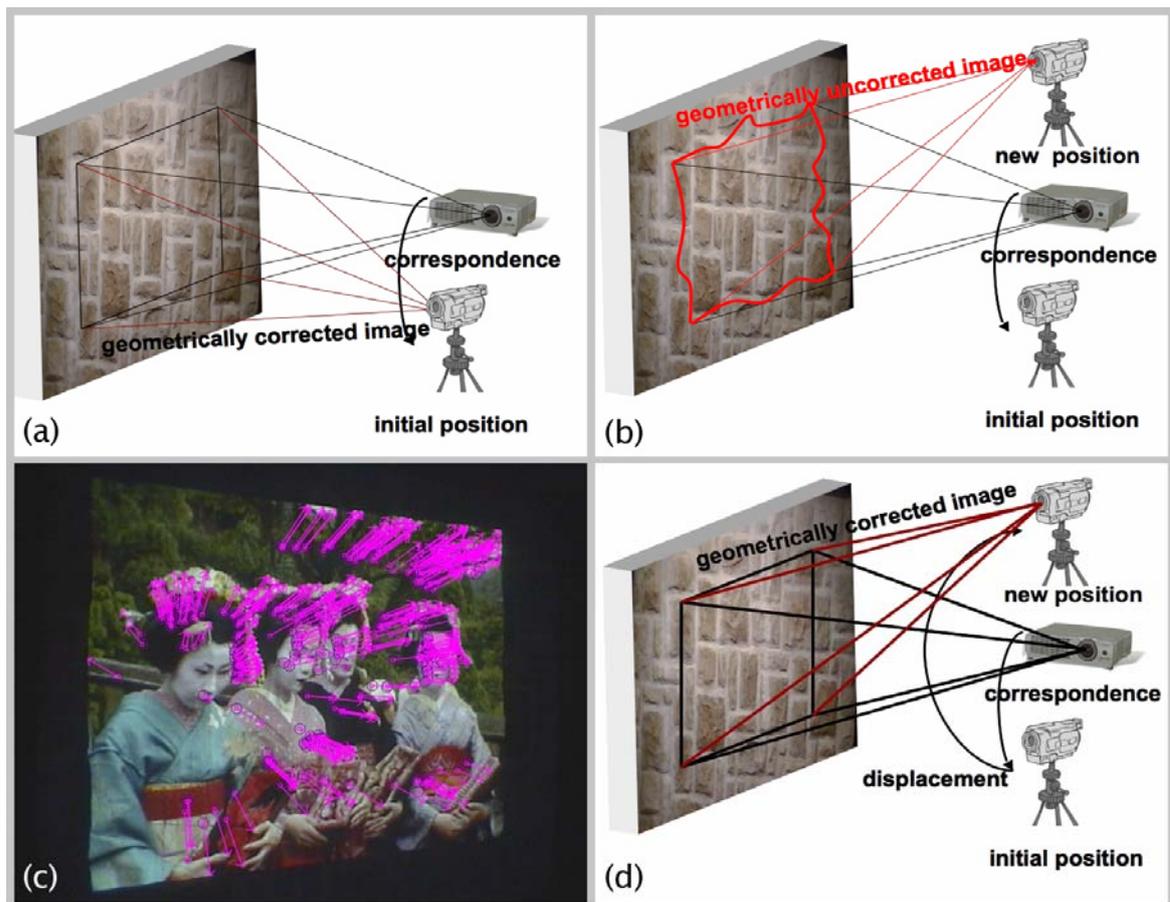


Figure 4: Passive calibration steps: correct image (a), geometric distortion for new camera position (b), computed optical flow vectors (c), geometrically corrected image based on optical flow analysis (d).

A second optical flow analysis between the image captured from C_i and the perspective corrected image computed via the homography matrix for C_i is performed next (line 7). While the first image contains geometric distortions due to the non-planar screen geometry, the latter one does not. Both images, however, approximate the same perspective distortion. Unreliable flow vectors are filtered out again. The determined optical flow vectors allows –once again–

determining the displacements for a number of discrete feature pixels between both images. We apply a pyramidal implementation of the Lucas Kanade Feature Tracker [Bou1999] for optical



Figure 5: Passive calibration results: ideal image (a) and close-up (e), uncorrected image (b) and close-up (f), corrected image (c) and close-up (g), image without radiometric compensation under environment light (d), visualized error maps for uncorrected (h) and corrected (i) case.

flow analysis. It enables to determine large pixel flows with sub pixel accuracy. Since this algorithm calculates the optical flow only for a sparse feature set, interpolation of the resulting flow vectors has to be applied to fill empty pixel positions. This makes it finally possible to establish the correspondences between each visible image pixel in C_i and C_0 . Since the mapping between C_0 and P is known, the mapping from C_i to P is implied. Only the optical flow calculations are carried out on the CPU. Both look-up tables are stored in texture maps and processed by a fragment shader that performs the actual pixel displacement mapping on the GPU in about 62ms.

Animations, movies or interactive renderings are paused during these steps to ensure equal image content for all computations. Thus, new images are continuously computed for perspective C_0 (i.e., rendered into the virtual canvas defined from C_0), warped into perspective C_i and finally transformed to the perspective of the projector P . For implementation reasons, the shader lookup operations are applied in a different order: $P \rightarrow C_0 \rightarrow C_i$.

3.4 Active Calibration for Large Perspective Changes

If the passive calibration method becomes too imprecise the accurate active calibration is triggered automatically to reset the mapping $C_0 \rightarrow P$. The quality of the passive correction can be determined by evaluating the amount of features with large eigenvalues and the number of valid flow vectors. If both drop under a pre-defined threshold the offline calibration will be triggered. The same active calibration technique as described in section 3.2 can be applied. The perspective change that is due to the camera movement, however, has to be considered.

As for the passive method, we try to simulate the perspective distortion between the previous C_0 and C_i by computing a homography matrix. The original image can then be warped from C_0 to C_i by multiplying every pixel in C_0 with this matrix. To determine the homography matrix we have to solve a linear equation system via least squares method for a minimum of nine sample points with known image coordinates in both perspectives. Since the camera-projector point correspondences in both cases are known (the original $C_0 \rightarrow P$ from the initial active calibration, and $C_i \rightarrow P$ from a second active calibration that is triggered if the passive calibration fails), the mapping from C_0 to C_i is implicitly given for every calibration point. As mentioned in section 3.3 the mapping from C_0 to C_i by using the homography matrix is only an approximation.

3.5 Radiometric Compensation

In addition to the geometric image distortion, the projected pixels' colors are blended by the reflectance of underlying surface pigments. This results in color artifacts if the surface has a non-uniform color and a texture. To overcome these artifacts, the pixels' colors are modified in such a way that their corresponding blended reflection on the surfaces approximates the original color. If the surface is Lambertian, this is view-independent, and a variety of radiometric compensation techniques can be applied [BiEm2005] [NaPe2003]. To compensate small view-dependent effects (such as light specular highlights), image-based techniques offer appropriate approximations [BiWe2005].

All of these techniques initially measure parameters, such as the environment light and projector contributions as well as the surface reflectance via structured light and camera feedback. We carry out our radiometric compensation computations for multiple projectors [BiEm2005] on a per-pixel basis directly the GPU – within the same fragment shader that performs the per-pixel displacement mapping (see section 3.3). Since we assume diffuse surfaces we can measure the required parameters during the initial calibration step (see sections 3.1 and 3.2) and map them (i.e., via $C_0 \rightarrow P$) to the image plane of the projector – which is static. Therefore, these parameters are constant all the time and a recalibration after camera movement is not necessary.

Note that all images that are captured for computing optical flow vectors (section 3.3) are radiometrically compensated to avoid color artifacts that would otherwise lead to an incorrect optical flow.

4 Summary and Future Work

In this paper we have presented a hybrid calibration technique for correcting view-dependent distortions that appear when projecting images onto geometrically and radiometrically complex surfaces. During camera movement (i.e., movement of the observer's target perspective), the optical flow of the displayed image is analyzed. If possible, a per-pixel warping of the image geometry is carried out on the fly – without projecting visible light patterns. However, if the optical flow is too unreliable, a fast active calibration is triggered automatically. Together with

an appropriate radiometric compensation, this allows perceiving undistorted images, videos, or interactively rendered (monoscopic or stereoscopic) content onto complex surfaces from arbitrary perspectives. We believe that for domains such as virtual reality and augmented reality this holds the potential of avoiding special projection screens and inflexible screen configurations. Arbitrary everyday surfaces can be used instead – even complex ones.

Since all parameters for radiometric compensation are mapped to the perspective of the projector, they become independent to the observer's perspective. Consequently, multi-projector techniques for radiometric compensation [BiEm2005] and multi-focal projection [BiEm2006] are supported.

Our future work will focus on replacing the active calibration step that yet displays visible patterns by techniques that project imperceptible patterns. This will lead to an invisible and continuous geometric and radiometric calibration process. The passive part of this process will ensure fast update rates – especially for small perspective changes. The active part still provides an accurate solution at slower rates. Both steps might also be parallelized to allow selecting the best solution available at a time.

Acknowledgments

This project is supported by the Deutsche Forschungsgemeinschaft (DFG) under contract number PE 1183/1-1.

References

- [BiWe2005] Bimber, O., Wetzstein, G., Emmerling, A., & Nitschke, C.: *Enabling View-Dependent Stereoscopic Projection in Real Environments*. Proc. of IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), pp. 14-23, Vienna, 2005.
- [BiEm2005] Bimber, O., Emmerling, A., Klemmer, T.: *Embedded Entertainment with Smart Projectors*. IEEE Computer, 38(1), pp. 56-63, 2005.
- [BiEm2006] Bimber, O., Emmerling, A.: *Multi-Focal Projection: A Multi-Projector Technique for Increasing Focal Depth*. IEEE Transactions on Visualization and Computer Graphics, pp. 658-666, Volume 12, 2006.
- [Bou1999] Bouguet, J.-Y.: *Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm*. Technical Report, Intel Corporation, Microprocessor Research Labs. OpenCV Documents. 1999.

- [CaKi1998] Caspi, D., Kiryati, N., Shamir, J.: *Range imaging with adaptive color structured light*, IEEE Transactions on Pattern analysis and machine intelligence, pp. 470-480, Volume 20, 1998.
- [CheSu2002] Chen, H., Sukthankar, R., Wallace, G.: *Scalable alignment of large-format multi-projector displays using camera homography trees*, Proc. of IEEE Visualization (IEEE Viz'02), pp. 339-346, 2002.
- [CoNa2004] Cotting, D., Naef, M., Gross, M., Fuchs, H.: *Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display*, IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR04), pp. 100-109, Arlington, 2004.
- [Gü2001] Gühring, J.: *Dense 3-d surface acquisition by structured light using off-the-shelf components*, Proc. Videometrics and Optical Methods for 3D Shape Measurement, pp. 220-231, Volume 4309, 2001.
- [NaPe2003] Nayar, S.K., Peri, H., Grossberg, M.D., Belhumeur, P.N.: *A Projection System with Radiometric Compensation for Screen Imperfections*, Proc. of International Workshop on Projector-Camera Systems, 2003.
- [PoAlt1982] Posdamer, J. L., Altschuler, M. D.: *Surface measurement by space-encoded projected beam systems*, Computer Graphics and Image Processing, pp. 1-17, vol. 18 issue 1, 1982.
- [RasBro1999] Raskar, R., Brown, M.S., Yang, R., Chen, W., Welch, G., Towles, H., Seales, B., & Fuchs, H.: *Multi-projector displays using camera-based registration*, Proc. of IEEE Visualization (IEEE Viz'99), pp. 161-168, 1999.
- [WaWü2005] Waschbüsch, M., Würmlin, S., Cotting, D., Sadlo, F., Gross, M.: *Scalable 3D Video of Dynamic Scenes*. The Visual Computer, Volume 21, pp. 629-638, 2005.
- [YanWe2001] Yang, R., Welch, G.: *Automatic and Continuous Projector Display Surface Calibration Using Every-Day Imagery*. Proc. of 9th International Conf. in Central Europe in Computer Graphics, Visualization, and Computer Vision, Plzen, 2001.
- [YanGo2001] Yang, R., Gotz, D., Hensley, J., Towles, H.: *PixelFlex: A Reconfigurable Multi-Projector Display System*. IEEE Visualizations, pp. 161-168, 2001.