

Mobile Augmented Reality based 3D Snapshots

Peter Keitler* Frieder Pankratz Björn Schwerdtfeger Daniel Pustka Wolf Rödiger
Gudrun Klinker
Technische Universität München
Christian Rauch Anup Chathoth
Vodafone Group Services GmbH – Vodafone Group R&D Germany
John Collomosse Yi-Zhe Song
University of Bath



Figure 1: Mobile AR viewer

ABSTRACT

We describe a mobile augmented reality application that is based on 3D snapshotting using multiple photographs. Optical square markers provide the anchor for reconstructed virtual objects in the scene. A novel approach based on pixel flow highly improves tracking performance. This dual tracking approach also allows for a new single-button user interface metaphor for moving virtual objects in the scene. The development of the AR viewer was accompanied by user studies confirming the chosen approach.

Index Terms: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles;

1 MOTIVATION

We present a mobile augmented reality (AR) platform based on user-generated content. The core idea is to enable a user of our system to generate a 3D model of arbitrary small or mid-sized objects, based on photographs taken with their mobile phone camera. Thereupon, another user can inspect the object integrated in their natural environment, e.g. their home, using our mobile AR viewer application, as shown in Figure 1. We refer to this capture and viewing process as “3D Snapshotting”.

The 3D reconstruction step is mainly based on methods described in [6, 5, 7, 4]. Basically, a textured mesh is formed over a 3D dense feature correspondence. In this work, the focus lays on

*keitler@in.tum.de

the AR viewer application. We investigated questions regarding the usability of the viewer, such as placement of objects or interaction with the scene.

2 MOBILE AR VIEWER REQUIREMENTS

For a first informal user study, we developed an AR viewer based solely on optical square marker tracking similar to [2]. The main focus of the study was to observe how the participants interact with the phone and the test application, to gather requirements for the tracking system and application interface (see Section 4). The participants were not told how the marker tracking system worked and what flaws it had. They were just informed that a virtual object would appear on top of the marker. The intention was to observe how the users would naturally interact with the system. Five participants had to align simple furniture objects with real objects, orient them in a certain direction or inspect them from different perspectives.

All participants moved the mobile phone too fast in the beginning. This resulted in strong motion blur causing the marker tracking to fail, an effect that was even amplified by the long shutter times of the integrated phone camera.

Another big problem occurred in conjunction with bigger objects such as tables and chairs. Since the field of view of our phone camera is rather small, the virtual objects occluded almost the complete display, making it difficult to inspect it in its environment without losing the marker. In case the marker was further away from the participant, the field of view wasn't as much of a problem, but the pose of the marker began to jitter, depending on its size, which the participants felt as very troublesome. In the end, the participants developed strategies to compensate for this, for example, standing completely still in front of the marker, interacting only via the touch manipulation of the AR viewer (see Section 4). Thus the application became less usable. From this first user study, we also learned that the participants preferred an inaccurate but stable pose over an unstable pose. During the user study, we also noticed that users typically do not translate the phone much but rather rotate it around axes lying in the wrist, elbow or hip. This offers the opportunity for simplified pose estimation when the marker is absent. Summarizing the results, the study suggests that optical square marker tracking alone is not enough to provide a satisfactory user experience.

3 THE TRACKER

Based on these initial experiences, a more elaborate tracking system was developed. Jitter was reduced by a double exponential smoothing predictor [3]. Furthermore, the optical square marker tracker was supplemented by the projection shift analysis (PSA) algorithm, originally designed to use the motion of the phone as an interaction technique, in order to compute the orientation of the phone [1]. The algorithm computes the horizontal and vertical shift between two consecutive frames. First, it calculates the sums of gray values for every row and column in the current image, resulting in two vectors of accumulated gray values, one for the horizontal and one

for the vertical shift. These vectors are compared with the vectors from the previous image to determine the shift of the current image. The optimal shift is found by minimizing the sum of squared differences (SSD) over all possible shifts. To increase the performance and stability of the algorithm, the SSD is only calculated for a predefined range of shifts. Tests suggest a range of 50% of the image size. To cope with the effects of motion blur, an image pyramid is used, starting with half the resolution of the original image, and iteratively dividing the resolution by two in case the SSD exceeds an upper limit. This can be implemented efficiently by just scaling down the row and column intensity vector results of the previous PSA execution instead.

As long as the marker is being tracked by the system, its pose is solely used. Once the marker is lost, the horizontal and vertical shift of the current image with respect to the previous image is used to estimate the orientation change of the mobile phone around its X and Y axes. The last known marker pose is incrementally updated using these values. Translation of the phone is assumed to be negligible compared to changes in orientation (as noted in our earlier study). The conversion from shift in the image to the rotation of the phone is done by using two factors, one for the X and one for the Y axis. They depend on various factors and are constantly being adjusted as long as the marker is still visible.

Having this, users don't need to keep the marker in the visible area all the time. They can interact more naturally with the application and their environment, even when the marker is lost. Another advantage is that the results of the pixel flow can be used for an alternative interaction method, as will be described next.

4 THE AUGMENTED REALITY VIEWER

The user interface of the AR viewer consists of a toolbar overlay. It contains buttons to load, select and erase virtual objects, to change settings, to toggle between view and object manipulation mode as well as several tools to arrange objects in space (move, rotate, lift and scale). They were identified in our first user study described in Section 2.

Drop shadows are rendered at the bottom of the objects to improve the sense of immersion. While being tracked, the square marker is highlighted with a green selection to provide visual feedback. During user interaction with one particular object all other objects are rendered transparently to emphasize the current selection. In object manipulation mode, two different interaction techniques are provided:

Touch Manipulation Once a function has been selected, the user has to drag a line on the screen with their finger. The relative movement on the screen is used to directly manipulate the attributes of the selected object, e.g. the position in the plane defined by the marker (move) or the height (lift), using a predefined factor for each function. These factors were defined using extensive testing.

Flow Manipulation Using the touch screen for user input requires both hands most of the time. Using the pixel flow provided by the tracking system, we can implement an interaction method for only one hand. First, the desired manipulation function is selected from the toolbar, preferably using the thumb. Then, instead of using the touch screen to, e.g. move an object, the user has to press a central button on the phone while moving the phone in the according direction(s). Again, the relative movement is used as input for the selected manipulation function, except that other conversion factors are used.

5 EVALUATION AND RESULTS

Based on the functionality described so far, we performed an informal user study with 8 subjects. Like in our first user study, the participants were not told at first how the tracking system worked. This time the participants had to place different objects onto other

objects, view an about 2 meters tall virtual advertising column and identify 3 posters that were placed around the top of it, thereby forcing them to use the pixel flow tracking.

The overall satisfaction when using the application increased drastically, partly because of the stabilized marker position, but mostly because the participants did not have to pay much attention to the marker and could freely move the phone without worrying about losing it. The participants were still standing in front of the marker, rotating the phone about an axis in the wrist, elbow or body, approving our assumption about the movements of the user. But with the freedom the new system granted, some users tried to view the top of the advertising column and walk around the object at the same time. The tracking system doesn't support such movements, so the users didn't get the result they expected. After explaining to them the restrictions of the new tracking system, the users quickly adapted to it. They either went around the marker, looked at the marker for a few moments to let the system register the new position or they used the interaction methods provided by the application to rotate the advertising column. Most of the participants didn't notice the drift at first or where not bothered by it. Only a few participants repeatedly checked whether the object was still at the marker position.

The user study showed that the flow manipulation is particularly suited for translating the object in the X/Y (move) and Z (lift) directions, but unusable for rotation since the required movement of the phone (rotating the phone around its Y axis) would make the object leave the field of view. Scaling the object using the flow manipulation is possible, but since it requires the phone to move, the point of view changes and so the users were not sure anymore to which size they wanted to scale the object since they lost their reference points. Even though the touch manipulation was at a much better stage of implementation at this time, the users preferred the flow manipulation for moving the object. Only when the given task required precision, the users switched to the touch manipulation.

ACKNOWLEDGEMENTS

This topic has been initiated at the Vodafone Group R&D academic flagship conference in November 2007. The research activities have been supported and supervised by Vodafone Group R&D.

REFERENCES

- [1] S. A. Drab and N. M. Artner. Motion detection as interaction technique for games & applications on mobile devices. In *Pervasive Mobile Interaction Devices (PERMID 2005) Workshop at the Pervasive 2005*, Munich, Germany, May 2005.
- [2] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, USA, October 1999.
- [3] J. J. LaViola Jr. An experiment comparing double exponential smoothing and kalman filter-based predictive tracking algorithms. In *VR '03: Proceedings of the IEEE Virtual Reality 2003*, page 283, Washington, DC, USA, 2003. IEEE Computer Society.
- [4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [5] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 1997.
- [6] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24:271–300, 1997.
- [7] N. S. University, N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *ACM Transactions on Graphics*, pages 835–846. Press, 2006.