OST Rift: Temporally Consistent Augmented Reality with a Consumer Optical See-Through Head-Mounted Display

Yuta Itoh* TU München
Jason Orlosky† Osaka University
Manuel Huber‡ TU München
Kiyoshi Kiyokawa§ Osaka University
Gudrun Klinker¶ TU München

ABSTRACT

We present an off-the-shelf, low-latency Optical See-through Head-Mounted Displays (OST-HMD) for Augmented Reality (AR). Temporally consistent visualization is crucial for realizing immersive AR experiences. This is challenging since it requires both accurate head-tracking and low-latency rendering of AR content. Building a system which meets both constraints usually requires experts on computer vision/graphics and expensive display hardware. This work demonstrates that such high spatio-temporal fidelity is achievable with commodity hardware available today. We build a custom OST-HMD system that consists of a virtual reality HMD, i.e., the Oculus Rift DK2, and half-mirror optics, and adapt the rendering pipeline in order to integrate the OST-HMD calibration framework. An evaluation with a user-perspective camera shows that the system achieves mean temporal error of <1 ms (95% reduction of the latency from naive, no-predictive rendering), and median spatial error <0.3° in the viewing angle with maximum error at most 1.0°.

Keywords: low-cost HMD, post-rendering, optical see-through

1 INTRODUCTION

Temporally consistent visualization is a key requirement for producing indistinguishable AR experiences with OST-HMDs [6]. This is a significant challenge since it requires both head-tracking in the real world and rendering of AR content with low-latency. Each of the two requirements has been difficult research issues in themselves, and combining them all together in a single device is a common goal in AR.

Fortunately, recent developments in low-cost closed HMDs for both virtual reality (VR) and gaming have taken us one step closer to this goal. The Oculus Rift DK2 realizes immersive, low-latency VR experiences at an affordable price. Its low-latency rendering pipeline designed for VR takes into account the user's head motion in the real world, and can display virtual images within 20 ms [4].

The advantage of this pipeline is that developers have access to its core rendering routine together with tracking data. A key technique used in the pipeline is an image warp technique that happens in post-rendering [5], which Oculus refers to as timewarp. This technique *compensates* for the delay between the user's current head pose and the rendered image by shifting the image in 2D based on the very last measurements from an inertial measurement unit (IMU) with a very high sampling rate (1,000Hz for DK2).

We demonstrate that a temporally and spatially consistent AR experience is possible with commodity devices. We combine the above image warp technology with a spatial calibration of OST-HMDs. We first build a custom OST-HMD which is based on the Oculus Rift DK2. We then modify the rendering pipeline of the HMD so that we can calibrate the display against a user-perspective camera. Since our OST-HMD system design assumes a common VR HMD architecture with a flat display panel and head tracking, our findings can easily be applied to other low-cost VR HMDs to produce a similar OST-HMD.

*e-mail: yuta.itoh@in.tum.de
†e-mail: orlosky@lab.ime.cmc.osaka-u.ac.jp
‡e-mail: huberma@in.tum.de
§e-mail: kiyo@ime.cmc.osaka-u.ac.jp
¶e-mail :klinker@in.tum.de

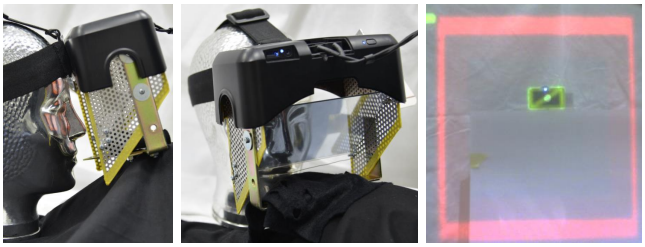

Figure 1: Our OST-Rift system. (Left/Middle) Side/Front view. A half-mirror installed bottom of the display housing reflects light from the panel to a user's view. (Right) A user perspective view through the HMD. A virtual 3D square is aligned on the cubic tracker camera.

In contrast to existing works that employ external hardware such as that of Zheng et al., who use a custom Digital Light Processing (DLP) projector [8], our system can be replicated with commodity hardware and is easier to build.

An evaluation shows that our system achieves the temporal error of less than 1 ms against the real world and the median spatial error less than 0.3deg for viewing angle (with the maximum error 1.0°) at the same time. We conclude with the limitations of the current system and discussion of how to improve the AR experience even further. Our main contributions include:
1) Realizing a temporally and spatially consistent AR on a low-cost OST-HMD system, which yields temporal/spatial error of <0.5ms/<0.3° in the viewing angle with at maximum error 1.0°.
2) Elaborating setup procedure to build and calibrate such a system with a commodity hardware.

2 SYSTEM OVERVIEW

Figure 1 and 2 show our prototype system and its image as seen by users. VR HMDs commonly split the display panel into two areas, one for the left eye and the other for the right, and insert magnifying lenses in front of the panel so that a user can increase stereo field of view. Since the panel occludes the view of the environment, these displays do not have OST capability.

To make an OST system, we place a half mirror and rotate the display 90° upwards, i.e., the display pannel looks at the ground after the rotation. We remove the lenses since the physical display is no longer near-eye. An apparent problem of this design is the reduced field of view and the accommodation distance [1] – though addressing these issues is not the focus of this paper.

As we rotate the display upwards, the user's head motion will not initially match the perceived image, which is seen as a plane floating in mid air from the user perspective view. We construct our virtual camera and convert 3D pose measurements of the real world into the eye coordinate system. One last thing we need to take care

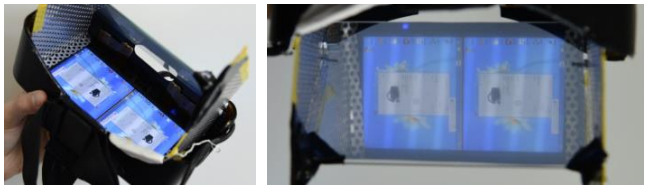

Figure 2: Our custom OST-HMD. (Left) Looking the display panel of the HMD. (Right) A user-perspective view through the OST-HMD.

of is the flipping effect due to the mirror. We handle all rendering as if there is no such effect, and the final stage of the rendering in the pixel shader, we simply flip over the x axis.

Hardware Figure 2 is our prototype OST-HMD system. Similar to the design by Dunn et al. [1], we kept the display and the sensor board rigidly attached to the outer housing so that we can still use the pose from the outside-in tracking system of the Oculus SDK. A half-mirror is rigidly attached to the HMD so that the mirror holds 45° against the display panel. The half mirror is a planer glass plate with reflective foil. The metal plates used to support the mirror are screwed onto the HMD housing.

Since we reflect light from the HMD screen through the mirror at 45°, the HMD housing was rotated 90° upwards relative to the user so that the screen looks at the mirror. The display panel is held in place by the inner housing with the lens mount, so we carefully removed the housing without moving the display panel, then glued the mirrored panel onto the outer housing. This process was necessary to fix the relative pose of the IMU on the mainboard of the HMD located behind the display panel to the IR LEDs on the outer housing. Otherwise, it is likely that the sensor fusion provided by the Oculus firmware would be degraded or not work due to the misalignement between the IMU and the LED constellation.

Software We used the Oculus SDK 0.5.1 to access the tracking data of the system and to build our visualization software on top of the rendering pipeline. Through the SDK, we can access the 6DoF pose data of the HMD housing from the DK2 firmware. We then modified their sample application code for visualization and the kernel code for customizing the timewarp pipeline. Since our system does not use convex lenses that introduce image distortion, we disabled predistortion rendering in the kernel code.

When integrating rendering and tracking software designed for VR to our AR setup, there were many pitfalls we needed to consider. The biggest problem was to adjust the FoV of the virtual camera so that AR rendering works properly, which was accomplished through our manual calibration. To get the correct head rotation, we also need to transform sensor measurements so that they are compatible with the camera. We built the transformation from the projection matrix we obtained during the calibration step.

3 Experiment

To evaluate the temporal and spatial errors of the proposed system, we employ a user-perspective camera to represent a user's eyeball. We calibrated the HMD screen with respect to the camera by a standard manual calibration. Note that the camera and OST Rift are not synchronized in the experiment.

We measure the latency of the system from the time it starts tracking until the time our eye sees the rendered AR content. Similar to the work by Steed [7], we record the periodical motion of a reference point in the real world in the HMD coordinate system while rendering an AR reference. We first rotate the entire hardware setup along the user-perspective camera while displaying AR content on the screen, followed by analysis of the recorded data captured by the camera during motion. We built a hardware setup that provides controllable and reproducible viewpoint motion. As a world reference, we attached a blue LED on top of the Oculus

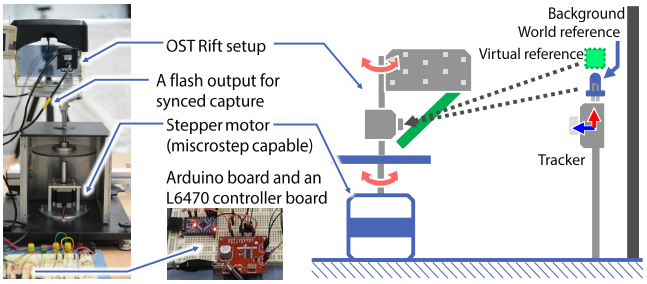

Figure 3: The rotation stage setup used in the experiment.

tracker. As a screen reference, we rendered a green cube in proximity to the tracker in the 3D space so that it appears above the LED from the camera perspective.

We analyze the delay of the virtual square reference relative (virtual) to the blue LED reference (real). We follow an approach similar to the time delay estimation for tracking data described in [2], which we adapted to our particular setup. We report that estimated time delays did not exceed 1 ms with the mean less than 0.5ms. Spatial registration error of the system was roughly 7.4 pixel in median with timewarp, which is roughly equivalent to 0.25°.

4 Conclusion and Future Work

We present a low-cost OST-HMD that achieves a mean latency of <1 ms and spatial registration error of <0.3° for viewing angle (with a max. error of 1.0°) at the same time. The display is based on the conventional VR technology, which makes its design both inexpensive and replicable. A possible future work is to integrate the automated calibration through eye tracking [3]. We hope the current work will provide easier access to low latency OST-HMDs and motivate others to build joyful AR experiences with their own hands.

5 Acknowledgments

The project received funding from the European Uinon under contract number: PITN-GA-2012- 316919-EDUSAFE.

References

[1] D. Dunn, A. State, K. Keller, and H. Fuchs. Converting commodity head-mounted displays for optical see-through augmented reality. Technical report, University of North Carolina at Chapel Hill, May 2015.

[2] M. Huber, M. Schlegel, and G. Klinker. Application of time-delay estimation to mixed reality multisensor tracking. *Journal of Virtual Reality and Broadcasting*, 11(2014):3, 2014.

[3] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3D eye localization. In *IEEE Symposium on 3D User Interfaces, 3DUI 2014, Minneapolis, MN, USA, March 29-30, 2014*, pages 75–82, 2014.

[4] S. LaValle. The latent power of prediction, https://developer.oculus.com/blog/the-latent-power-of-prediction/, OCULUS VR, LLC.

[5] W. R. Mark, L. McMillan, and G. Bishop. Post-rendering 3d warping. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 7–ff. ACM, 1997.

[6] M. Mine. Characterization of end-to-end delays in head-mounted display systems. *The University of North Carolina at Chapel Hill, TR93-001*, 1993.

[7] A. Steed. A simple method for estimating the latency of interactive, real-time graphics simulations. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 123–129. ACM, 2008.

[8] F. Zheng, T. Whitted, A. Lastra, P. Lincoln, A. State, A. Maimone, and H. Fuchs. Minimizing latency for augmented reality displays: Frames considered harmful. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 195–200. IEEE, 2014.