Headset Removal for Virtual and Mixed Reality

Christian Frueh
Google Research
frueh@google.com

Avneesh Sud
Google Research
avneesh@google.com

Vivek Kwatra
Google Research
kwatra@google.com

Figure 1: Mixed Reality Overview: VR user captured in front of green-screen (A), blended with her virtual environment (from *Google Tilt Brush*) (B), to generate the MR output: Traditional MR output (C) has the user face occluded, while our result (D) reveals the face. The headset is deliberately rendered translucent instead of being completely removed.

ABSTRACT

Virtual Reality (VR) has advanced significantly in recent years and allows users to explore novel environments (both real and imaginary), play games, and engage with media in a way that is unprecedentedly immersive. However, compared to physical reality, sharing these experiences is difficult because the user's virtual environment is not easily observable from the outside and the user's face is partly occluded by the VR headset. Mixed Reality (MR) is a medium that alleviates some of this disconnect by sharing the virtual context of a VR user in a flat video format that can be consumed by an audience to get a feel for the user's experience.

Even though MR allows audiences to connect actions of the VR user with their virtual environment, empathizing with them is difficult because their face is hidden by the headset. We present a solution to address this problem by virtually *removing* the headset and revealing the face underneath it using a combination of 3D vision, machine learning and graphics techniques. We have integrated our headset removal approach with Mixed Reality, and demonstrate results on several VR games and experiences.

CCS CONCEPTS

•**Computing methodologies → Mixed / augmented reality; Virtual reality;**

KEYWORDS

Mixed reality, virtual reality, headset removal, facial synthesis

1 APPROACH

Creating Mixed Reality videos [Gartner 2016] requires a specialized, calibrated setup consisting of an external camera attached to a VR controller and time-synced with the VR headset. The camera captures the VR user in front of a green screen, which allows compositing a cutout of the user into the virtual world, using headset telemetry to correctly situate the real and virtual elements in appropriate layers. However, the occluding headset masks the identity of the user, blocks eye gaze, and renders facial expressions and other non-verbal cues incomplete or ineffective. This presents a significant hurdle to a fully engaging experience.

We enhance Mixed Reality by augmenting it with our headset removal technique that creates an illusion of revealing the user's face (Figure 1). It does so by placing a personalized face model of the user behind the headset in 3D, and blending it so as to create a see-through effect in real-time. This is done in three steps.

1.1 Dynamic 3D Face Model Capture

First, we capture a personalized dynamic face model of the user in an offline process, during which the user sits in front of a calibrated setup consisting of a color+depth camera and a monitor, and follows a marker on the monitor with their eyes. We use this one-time procedure—which typically takes less than a minute—to create a 3D model of the user's face, and learn a database that maps appearance images (or textures) to different eye-gaze directions and blinks.

This gaze database allows us to dynamically change the appearance of the face during synthesis and generate any desired eye-gaze, thus making the synthesized face look natural and alive.

1.2 Automatic Calibration and Alignment

Secondly, compositing the human face into the virtual world requires solving two geometric alignment problems:

Calibration: We first estimate the calibration between the external camera and the VR headset (e.g. the *HTC Vive* used in our MR setup). Accuracy is important since any errors therein would manifest themselves as an unacceptable misalignment between the 3D model and the face in the camera stream. Existing mixed

reality calibration techniques involve significant manual intervention [Gartner 2016] and are done in multiple steps: first estimating the camera intrinsics like field-of-view, and then computing the extrinsic transformation between the camera and VR controllers. We simplify the process by adding a marker to the front of the headset, which allows computing the calibration parameters automatically from game play data—the marker is removed virtually during the rendering phase by *inpainting* it from surrounding headset pixels.

Face alignment: To render the virtual face, we need to align the 3D face model with the visible portion of the face in the camera stream, so that they blend seamlessly with each other. A reasonable proxy to this alignment is to position the face model just behind the headset, where the user's face rests during the VR session. This positioning is estimated based on the geometry and coordinate system of the headset. The calibration computed above is theoretically sufficient to track the headset in the camera view, but in practice there may be errors due to drift or jitter in the Vive tracking. Hence, we further refine the tracking (continuously in every frame) by rendering a virtual model of the headset from the camera viewpoint, and using silhouette matching to align it with the camera frame.

1.3 Compositing and Rendering

The last step involves compositing the aligned 3D face model with the live camera stream, which is subsequently merged with the virtual elements to create the final MR video. We identify the part of the face model likely to correspond to the occluded face regions, and then render it over the camera stream to fill in the missing information. To account for lighting changes between gaze database acquisition and run-time, we apply color correction and feathering so that the synthesized face region matches the rest of the face.

Dynamic gaze synthesis: To reproduce the true eye-gaze of the user, we use a Vive headset modified by SMI to incorporate eye-tracking technology. Images from the eye tracker lack sufficient detail to directly reproduce the occluded face region, but are well-suited to provide fine-grained gaze information. Using the live gaze data from the tracker, we synthesize a face proxy that accurately depicts the user's attention and blinks. We do so by searching the pre-built gaze database, at runtime, for face images that correspond to the live gaze state, while using interpolation and blending to respect aesthetic considerations like temporal smoothness.

Translucent rendering: Humans have high perceptual sensitivity to faces, and even small imperfections in synthesized faces can feel unnatural and distracting, a phenomenon known as the *uncanny valley*. To mitigate this problem, instead of removing the headset completely, we choose a user experience that conveys a 'scuba mask effect' by compositing the color-corrected face proxy with a translucent headset. Reminding the viewer of the presence of the headset helps avoid the uncanny valley and also makes our algorithms robust to small errors in misalignment and color correction.

2 RESULTS AND DISCUSSION

We have used our headset removal technology to enhance Mixed Reality, allowing it to convey not only the user's interaction with VR but also reveal their face in a natural and convincing fashion. Figure 2 demonstrates results on an artist using *Google Tilt Brush.*

Figure 3 shows another MR output from VR game-play, with before-and-after comparison. For more results, refer to our blog post [Kwatra et al. 2017]. Our tech can be made available on-request to creators at select YouTube Spaces (contact: tsmall@google.com).

Figure 2: Headset Removal in Mixed Reality

Figure 3: Before (left) and after (right) headset removal. Left image also shows the marker used for tracking.

Facial modeling and synthesis for VR is a nascent area of research. Recent work has explored advanced techniques for transferring gaze and expressions to target videos [Thies et al. 2016] and headset removal by reproducing expressions based on visual clustering and prediction [Burgos-Artizzu et al. 2015]. In contrast, our approach mimics true eye-gaze of the user, and is a practical end-to-end solution for headset removal, fully integrated with Mixed Reality.

Beyond MR, headset removal is poised to enhance communication and social interaction in VR with diverse applications like 3D video conferencing, multiplayer gaming, and co-exploration. We expect that going from a completely blank headset to being able to see, with photographic realism, the faces of fellow VR users will be a big leap forward in the VR world.

REFERENCES

Xavier P. Burgos-Artizzu, Julien Fleureau, Olivier Dumas, Thierry Tapie, François LeClerc, and Nicolas Mollet. 2015. Real-time Expression-sensitive HMD Face Reconstruction. In *SIGGRAPH Asia 2015 Technical Briefs*. ACM.

Kert Gartner. 2016. Making High Quality Mixed Reality VR Trailers and Videos. (2016). http://www.kertgartner.com/making-mixed-reality-vr-trailers-and-videos

Vivek Kwatra, Christian Frueh, and Avneesh Sud. 2017. Headset "Removal" for Virtual and Mixed Reality. (2017). https://research.googleblog.com/2017/02/headset-removal-for-virtual-and-mixed.html

Justus Thies, Michael Zollöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2016. FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality. *arXiv preprint arXiv:1610.03151* (2016).