High frequency 3D reconstruction of player hands for Virtual Reality

Karol B. Gasinski
Independent Contributor, kuktus@gmail.com

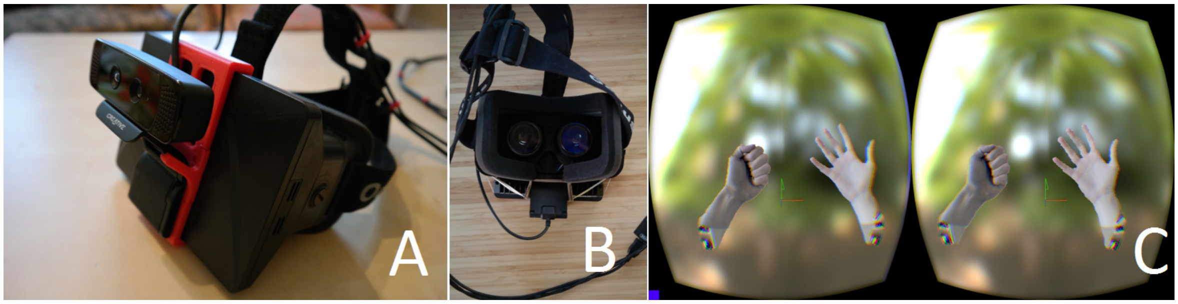

Figure 1: *A) First prototype assembled from Oculus DK1 and CreativeSenz3D depth camera. B) Second prototype assembled using Oculus DK2 and RealSense DK. C) Reconstructed view of player hands.*

1 Abstract

The method decouples reconstructing a hand's mesh from palm/skeleton detection, allowing for sub-millisecond processing times required by games rendering at 90-120Hz frequencies. The reconstructed mesh has the real texture and shape of the user's hands, drastically increasing player immersion, and can be illuminated to integrate seamlessly into the scene.

2 Introduction

While problems related to tracking and rendering for Virtual Reality are mostly solved, handling user interaction is still an issue. Complicated keyboard interfaces need to be revisited in favor of joysticks and gamepads. Yet the most intuitive solution - using your own hands is still not handled properly.

Current hand-detection solutions analyze incoming depth-buffer frames with sophisticated algorithms to detect palms and fingers, and extract the skeletal position and orientation of their bones [Bray, et al. 2004]. That skeletal information is used to render the user's hands as bone-animated meshes, as well as for collision detection. Drawbacks of such algorithms include that they tend to loose the tracked skeleton when fingers become occluded (e.g., by each other or by an object held in the hand). These algorithms are also too computationally expensive to be used in shipping a VR game, which requires a constant 90-120Hz framerate for rendering to preserve the immersive experience.

3 Implementation

The proposed method reuses information from incoming depth-buffer frames for mesh reconstruction in screen space. The process is divided into two parts. First, the algorithm attempts to find up to two depth samples that are part of the user's hands, using the sensor's orientation on the HMD. It then performs a two-pass filter operation that categorizes depth-buffer samples that are not part of detected hands. This step is performed on-the-fly during copying of the buffer from the depth camera to a memory-mapped texture. The second step reconstructs the mesh in screen space by analyzing incoming depth and color textures. The resulting mesh can be further integrated using in-game shadowing and lightning calculations to increase user presence in the scene.

The algorithm is highly parallelized on both CPUs and GPUs, and can be optimized uniquely for either computing environments. Palm/skeleton detection and tracking can still be performed for collision detection, but is not required for visualization, maintaining the 90-120Hz rendering frequency for comfortable immersive experiences. Additionally, the proposed method also allows proper visualization of the user's hands even when the skeletal information cannot be retrieved due to occlusion.

4 Prototype

The proposed algorithm was tested using a HW prototype assembled from an Oculus DK1 Head Mounted Display and CreativeSenz3D depth camera (Figure 1A). With the progress in the field of VR and depth sensors, the aforementioned HW setup was replaced by an Oculus DK2 and Real Sense DK (Figure 1B). To test the algorithm in the most demanding conditions, all measurements were taken using the latest HW prototype, since Real Sense DK is currently the most advanced depth camera available in the market, with a depth sensor resolution of 640x480 samples.

5 References

BRAY , M., KOLLER-MEIER, E., MULLER, P., VAN GOOL, L. AND SCHRAUDOLPH, N. N. 2004. 3D hand tracking by rapid stochastic gradient descent using a skinning model. Proc. of the 1st European Conf. on Visual Media Production, London, England, March 2004, pp. 59-68.