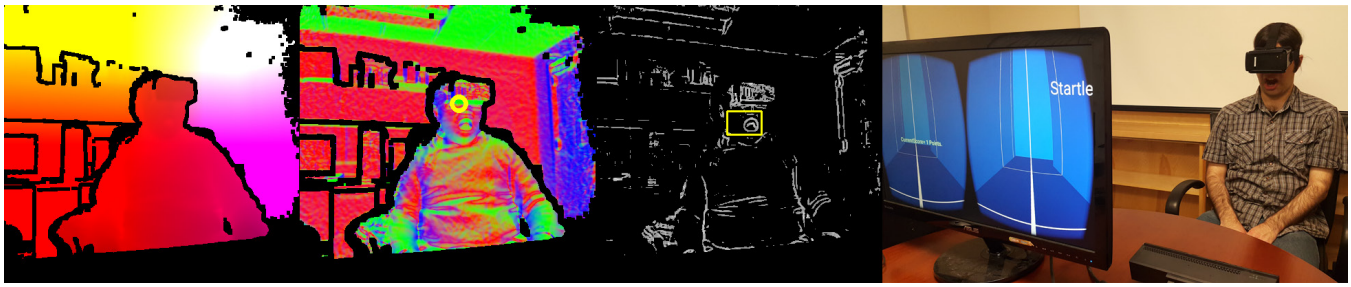
Mouth Gesture based Emotion Awareness and Interaction in Virtual Reality

Xing Zhang, Umur A Ciftci, and Lijun Yin*
SUNY-Binghamton University



(a) (b) (c) (d)

Figure 1: A user wears the VR headset and presents in front of the RGBD camera. (a) Encode 3D vertices from the camera space into the texture space. (b) Generate the normal map from the position map, with yellow circle indicating the automatically detected head position. (c) Extract contour from the normal map. The best matching expression is detected within the yellow rectangle region. (d) The detected expressions are used as the control for VR applications.

1 Introduction

In recent years, Virtual Reality (VR) has become a new media to provide users an immersive experience. Events happening in the VR connect closer to our emotions as compared to other interfaces. The emotion variations are reflected as our facial expressions. However, the current VR systems concentrate on "giving" information to the user, yet ignore "receiving" emotional status from the user, while this information definitely contributes to the media content rating and the user experience. On the other hand, traditional controllers become difficult to use due to the obscured view point. Hand and head gesture based control is an option [Cruz-Neira et al. 1993]. However, certain sensor devices need to be worn to assure control accuracy and users are easy to feel tired. Although face tracking achieves accurate result in both 2D and 3D scenarios, the current state-of-the-art systems cannot work when half of the face is occluded by the VR headset because the shape model is trained by data from the whole face.

To sense the emotional expression and receive natural input for VR systems, we propose to explore the merely exposed, but highly informative facial region - mouth. In this paper, we present a mouth contour extraction algorithm based on surface normal derived from 3D point cloud. The contour feature can clearly represent the mouth gestures in different facial expressions. The posed mouth gestures can also be the control in VR naturally.

2 Technical Approach

Motivated by the hatch rendering approach [Praun et al. 2001], we propose to use geometric edges to represent the mouth gestures. We begin with 3D point cloud, and use multiple image- filtering algorithms to extract the feature. The method is invariant to illumination and skin tone change. As shown in Figure 1(a), we encode the 3D position into the RGB channels. Gaussian smoothing is then applied to reduce noise and enhance feature in the position map. We calculate the normal vector of each pixel by directly reading the values of its eight neighbors and itself in the position map (Figure 1(b)). The normal map is accessed to extract edges by the formula $\sum_{i=1}^{K}(1 - N_0 \cdot N_i) > t$. We evaluate the sharpness of the edge pixel by the angle between its normal vector N_0 and its neighbors' normal vectors N_i as shown in the formula. By removing the non-edge pixels, which give values no more than the threshold t, we preserve the important contour features in the point cloud data (Figure 1(c)). The algorithm is general to any point cloud sources. In this research we apply the algorithm to the Kinect V2 point cloud, and implement it with CUDA in real time.

Rather than finding the head based on the facial features, we do it by utilizing the skeleton tracking result. This method is robust to the face occlusion condition. Then the search region of mouth can be defined as the yellow rectangle in Figure 1(c). Several different mouth gestures templates (currently three kinds including Happy, Startle, and Dislike) are defined. We search in the region with templates to locate the best matching and identify the mouth gesture. If all templates fail to match, indicated by the small correspondence factor, we treat this as the neutral expression.

3 Application and Future Work

The mouth gesture recognition pipeline is invariant to other facial features except mouth. At present it can detect four expressions including neutral, startled, happy, and dislike, whether the user wears VR headset or not. We have also developed a VR game (Figure 1(d)) where the player uses the mouth gestures to interact with the displayed scenes very easily. Our future work will apply machine learning approaches to support more mouth gestures.

Reference

CRUZ-NEIRA, C., SANDIN, D. J., & DEFANTI, T. A. (1993). Surround- screen projection-based virtual reality: the design and implementation of the CAVE. *SIGGRAPH 1993 Proceedings.*

PRAUN, E., HOPPE, H., WEBB, M., & FINKELSTEIN, A. (2001, August). Real-time Hatching. *SIGGRAPH 2001 Proceedings.*

*e-mail: {xzhang7, ncilsal2, lyin}@binghamton.edu
We thank the NSF for support under grant CNS-1205664.

SIGGRAPH 2015 Posters, August 09 – 13, 2015, Los Angeles, CA.
ACM 978-1-4503-3632-1/15/08.
http://dx.doi.org/10.1145/2787626.2787635