AR in Hand: Egocentric Palm Pose Tracking and Gesture Recognition for Augmented Reality Applications

Hui Liang[1], Junsong Yuan[2], Daniel Thalmann[1], Nadia Magnenat-Thalmann[1]
[1]Institute for Media Innovation, [2]School of Electrical and Electronics Engineering
50 Nanyang Avenue, Nanyang Technological University, Singapore
{lianghui,jsyuan,danielthalmann,nadiathalmann}@ntu.edu.sg

ABSTRACT

Wearable devices such as Microsoft Hololens and Google glass are highly popular in recent years. As traditional input hardware is difficult to use on such platforms, vision-based hand pose tracking and gesture control techniques are more suitable alternatives. This demo shows the possibility to interact with 3D contents with bare hands on wearable devices by two Augmented Reality applications, including virtual teapot manipulation and fountain animation in hand. Technically, we use a head-mounted depth camera to capture the RGB-D images from egocentric view, and adopt the random forest to regress for the palm pose and classify the hand gesture simultaneously via a spatial-voting framework. The predicted pose and gesture are used to render the 3D virtual objects, which are overlaid onto the hand region in input RGB images with camera calibration parameters for seamless virtual and real scene synthesis.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Depth cues, Tracking*

Keywords

Augmented Reality, Gesture Recognition, Palm Pose Estimation.

1. INTRODUCTION

Augmented Reality (AR) applications are now getting increasingly popular with the wearable devices such as Microsoft HoloLens and Google glass, which provide users enhanced reality by synthesizing real visual cues with virtual graphics. As portability matters much for wearable devices, the traditional input tools like mouse and keyboard are very inconvenient to carry and use on such platforms. By comparison, vision-based hand gesture control is more suitable, since no separate hardware is needed and it is more straightforward for users to directly use their bare hands and fingers to fulfill the various control and manipulative tasks. Despite that there exist several AR systems that are built upon vision-based hand tracking and gesture recognition [6, 2, 3], they still have limitations, *e.g.* sensitivity to hand posture variations or working in quite limited viewpoints.

In this demo we present an AR system to allow users to manipulate virtual objects freely with their bare hands based on our joint palm pose tracking and gesture recognition algorithm, so that the virtual objects are visually put upon the palm for arbitrary poses in the RGB images and the hand gestures are used to change the object appearance, *e.g.*, color. The framework of our system is shown in Fig. 1, in which we use a head-mounted SoftKinetic DS325 sensor to capture both the RGB and depth images of user's hand from egocentric view and predict the 6-DOF palm pose and recognize hand gesture simultaneously from the input depth images. The depth and color cameras of the SoftKinetic sensor are calibrated in advance to transform the 3D palm position and rotation recovered from the depth images to the coordinate

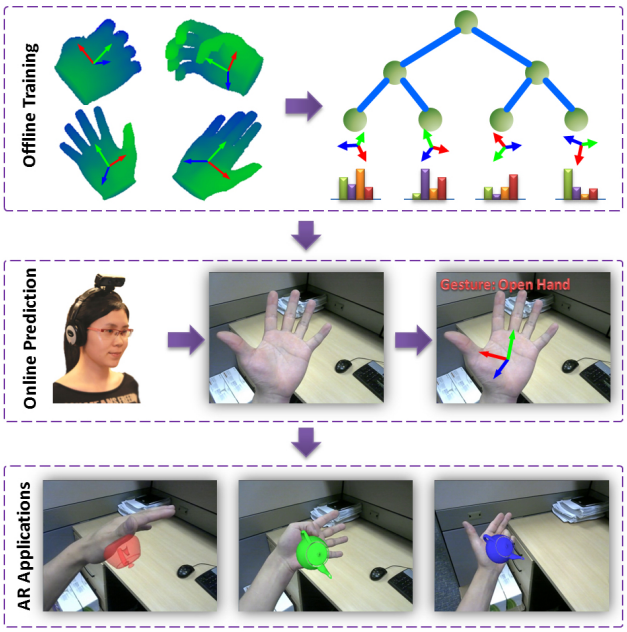

Figure 1: Framework of our AR system via egocentric palm pose tracking and gesture recognition.

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807972.

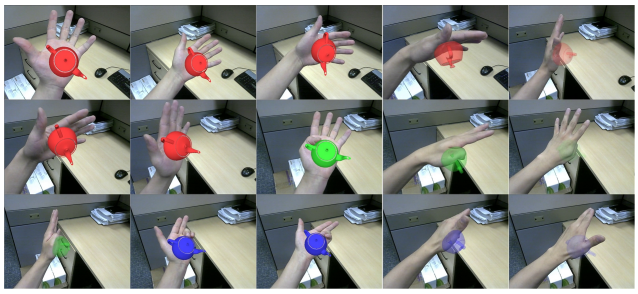

Figure 2: Teapot inspection and color selection.

Figure 3: Fountain animation in hand.

system centered at the color camera. The virtual objects are then rendered according to the transformed palm pose and hand gesture, and projected onto the image plane with OpenGL, which is then overlaid on the RGB image.

Technically, a random forest [4] is trained offline for joint regression of palm pose and classification of hand gestures so that they can be predicted together. Following previous work on spatial-voting based pose estimation and gesture recognition [7, 5], the random forest is learned to map the local features of spatially-distributed voting pixels to the probabilistic votes for either palm pose or gesture class. During online testing, a set of voting pixels are randomly sampled within the hand region in the depth image, which then retrieve their votes for palm pose and hand gesture with the random forest. The final pose and gesture predictions are obtained by MAP inference via a linearly weighting model of the per-pixel pose and gesture votes. Moreover, as such a linear weighting model with uniform weights tends to produce ambiguous predictions, we learn an optimal weighting model to fuse the per-pixel votes, which proves quite robust against noisy inputs. A Kalman filter is used to further smooth palm pose prediction in successive frames.

2. APPLICATION SCENARIOS

The goal of this demo is to showcase two AR applications built upon our joint palm pose tracking and hand gesture recognition algorithms, which allow users to interact with virtual objects with their bare hands. The system recognizes three hand gestures to change the object appearance and predicts unconstrained 6D palm pose to change the viewpoint for object inspection. The resolution of the depth images is 320×240. The program is coded in C++/OpenCV, and tested on a PC with Intel i5 750 CPU and 4G RAM. The average time cost to process one frame is less than 40ms, which is sufficient for real-time interaction.

In the first application users can manipulate and inspect a static teapot from different perspectives and use different gestures to change its color. Particularly, to get realistic visual feedback, we define a visibility term for the teapot based on hand rotation angles to reflect hand-object occlusion, which is implemented via controlling the transparency effect in OpenGL with the roll angle of the palm. That is, the teapot is fully opaque when the palm is facing the camera, and becomes gradually transparent when it rotates backwards. The user interface is shown in Fig. 2.

In the second application users can view a fountain animation in their hands and use different gestures to change the fountain style, as illustrated in Fig. 3. For fountain simulation we adopt the algorithm in [1], which models the fountain as a discrete set of water drops and changes their velocity by simulated gravity force. The different fountain styles are simulated by setting different numbers and initial velocities of water drops. The palm position and orientation are used as the origin to simulate the gravity force, and the water drops thus visually fall towards the palm plane.

3. CONCLUSION

This demo presents two AR applications built upon our unified 6-DOF palm pose tracking and gesture recognition algorithm, including static teapot manipulation and fountain animation in hand. Based on the recovered palm pose and camera calibration parameters, the virtual objects are rendered and overlaid onto the hand in the input RGB images, which provides seamless virtual and real scene synthesis. The program runs in real-time, and is particularly suitable for interaction on wearable devices.

4. ACKNOWLEDGMENT

This research, which is carried out at BeingThere Centre, is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

5. REFERENCES

[1] Fountain simulation. http://www.codecolony.de.

[2] O. Akman, R. Poelman, W. Caarls, and P. Jonker. Multi-cue hand detection and tracking for a head-mounted augmented reality system. *Machine Vision and Applications*, 24(5):931–946, July 2013.

[3] M. Asad and G. Slabaugh. Hand orientation regression using random forest for augmented reality. In *Augmented and Virtual Reality*, pages 159–174, 2014.

[4] L. Breiman. Random forests. *Mach. Learning*, 45(1):5–32, 2001.

[5] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun. Hand pose estimation and hand shape classification using multi-layered randomized decision forests. In *European Conf. Computer Vision*, 2012.

[6] T. Lee and T. Hollerer. Handy ar: Markerless inspection of augmented reality objects using fingertip tracking. In *IEEE Int'l Symposium on Wearable Computers*, pages 83–90, 2007.

[7] C. Xu and L. Cheng. Efficient hand pose estimation from a single depth image. In *IEEE Int'l Conf. Computer Vision*, 2013.