Physical Manipulation of Virtual Holograms Using Proxies

Alexa F. Siu
Stanford University
Department of Mechanical Engineering
afsiu@stanford.edu

Michael A. Lin
Stanford University
Department of Mechanical Engineering
mlinyang@stanford.edu

Note: Michael Lin is using this project for both EE267 and CS231A final projects.

1. Introduction

Recent technological advances in head-mounted optical-see-through displays (i.e. Microsoft HoloLens), have bridged the gap between the digital and the physical world. These mixed reality devices combine world tracking and head-mounted displays to show 3D image content that is anchored to the physical world. However, these efforts have mostly just increased use of humans visual display channels while still relying on the traditional point-and-click input for manipulating digital content.

In this project, we hope to bridge the gap between the input controls (i.e. pointer, mouse, keyboard) and the graphical output (i.e. the display) such that users can leverage the rich affordances of physical objects. Billinghurst et. al. defined Tangible AR interfaces as those in which: 1) virtual objects are registered to a real physical object (tangible) and 2) tangibles are used to manipulate the virtual objects [1]. Towards this vision of seamless tangible AR, we propose a novel way of interacting with holograms embedded in the real world. We propose using physical objects that are typically found in our surroundings (e.g. a smartphone, an eraser, a water bottle, a coffee mug etc) as handles or physical proxies to which holograms can virtually attach. Instead of using abstracted controls, the user would be able to manipulate the hologram just as they would manipulate any other physical object.

Beyond just moving physical objects, we also propose a system capable of understanding how a user interacts with the hologram. This enables holograms to act not only as visuals in our surroundings but to also react to user input, e.g. if a user is holding a non-empty virtual water bottle and then tilts it to pour onto a glass, water would pour out. On the other hand, if the user where to tilt the bottle with the bottleneck covered with the other hand, then water would not pour out. Holograms and their proxies can be seen as a new form of input controls to the 3D virtual content.

The ideal system would be able to track any arbitrary object chosen by the designer as a proxy. It would then overlay a hologram at the grasping location. For example, if the hologram were a lamp and the proxy was a water bottle, the system would align and overlay the lamps stem to the water bottle as this is considered the human choice of the lamps graspable region. This requires a system capable of real-time tracking from RGBD data. We propose using a probabilistic model-based tracking algorithm as detailed by Yuheng et al. [2]. This algorithm enables real-time tracking on a CPU with enhanced speed if also parallelizing computation on the GPU; but requires a model of the tracking object. We want to implement this algorithm to also generate models based on input dimensions.

The second problem we need to address is how to determine user interaction with the hologram and proxy. We limit interactions to hand actions that occur close to or on the physical proxy. With this constraint, we can segment the hand in the video frame based on a computed skin color histogram in a bounding box containing the hologram. Similar to how Yuheng et al. implemented the real-time tracking, we will take a probabilistic approach using Hidden Markov Model (HMM) to determine whether there was a user intent for an action. The motivation behind this approach is that working with the Kinect sensor may be noisy so we want to filter states using our and carefully take into account uncertainty. MDP is known to be a good approach for this.

2. Technical Approach

We divide our work timeline into three phases: real-time tracking, MDP for state estimation, and system integration.

2.1. Identifying and tracking the physical proxy

For tracking we use RGBD data as input and estimate the object pose at each frame. We follow the probabilistic model algorithm proposed by Yuheng et. al [3, 2]. This algorithm is set as a minimization problem solved using the gradient-descent method of Levenberg-Marquardt. The known model of the object being tracked must be known a priori as a signed distance function (SDF). At each frame the goal is to maximize the posterior probability of the pose

Figure 1. Milestones achieved. Sequence of screenshots from video showing how we can track a can based on a cylinder model

given the current depth image and 3D shape. This probability is calculated at a per-pixel likelihood and its energy function is defined as the sum of log-likelihoods of this probability over all pixels. Finally, the gradient that is iteratively minimized, is given by the change in energy over the change in pose.

2.2. MDP State Estimation

We first use HSV color histogram to segment anything around our object of interest that is of skin color. Then we take this blob of RGB and depth values as our MDP observations. We still need to work out the details of how we would define the state and how they inform the output, but generally the model would be as shown in Fig. 2. H.X represents the state being estimate at time X, and O.X represent the outcome at time X. The outcome only has two possibilities, there was or there was no action intent. The rest of the details still need to be worked out.

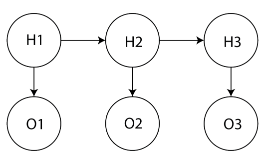

Figure 2. Hidden Markov Model

2.3. Kinect-Hololens Integration

Despite the Hololens having a depth camera, currently developers are only given access to its RGB camera. Since we require depth data for tracking, we plan to use data from an anchored Kinect for the tracking and then simply communicating to the Hololens where the hologram should be rendered in world coordinates. This will require some additional transformation of the coordinates from each camera frame to the world frame. Communication between the Kinect computer and HoloLens occurs through the network using TCP sockets.

3. Milestones Achieved

We have successfully replicated results from the algorithm proposed by Yuheng et. al for tracking a can but using the Kinect v2 instead of v1 like in the referenced paper [3, 2]. A series of screenshots from a video are shown in Figure 1. In this example, we use a coke can as physical proxy and a virtual blue cylinder is overlaid on the tracked object. We are working on varying the input models capable of being tracked. We would like to measure performance when simple shape primitives are used as models that approximate more complicated objects. For example, if the user wants to use a backpack as proxy, they could generate a generic cuboid as model with input dimensions.

4. Remaining Milestones

1. **[05/28] Refine algorithm for tracking the physical proxy**

   05/23 Generalize to enable tracking of cylinder, cube, and spheres of arbitrary size as determined by use

   05/26 Integrate with script for communicating through sockets with the Hololens.

   05/28 Gather results from tracking various shape proxies.

2. **[06/03] Implement algorithm for tracking user action intent.**

   5/28 Implement hand segmentation.

   5/30 Experiment with HMM model.

   06/03 Have a working implementation of the action intent estimation.

3. **[06/09] Project Presentations**

   06/03 Start preparing demo/slides

   06/08 Practice demo

4. **[06/12] Project Final Report Due: 11:59PM**

   06/03 Start writing draft

   06/10 Final draft

   06/12 Revise and refine

References

[1] M. Billinghurst, R. Grasset, H. Seichter, and A. Dünser. Towards ambient augmented reality with tangible interfaces. *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, pages 387–396, 2009.

[2] C. Y. Ren, V. Prisacariu, O. Kaehler, I. Reid, and D. Murray. 3d tracking of multiple objects with identical appearance using rgb-d input. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 47–54. IEEE, 2014.

[3] C. Yuheng Ren, V. Prisacariu, D. Murray, and I. Reid. Star3d: Simultaneous tracking and reconstruction of 3d objects using rgb-d data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1561–1568, 2013.