3D Tracking of Human Hands in Interaction with Unknown Objects

Paschalis Panteleris[1]
padeler@ics.forth.gr
Nikolaos Kyriazis[1]
kyriazis@ics.forth.gr
Antonis A. Argyros[21]
argyros@ics.forth.gr

[1] Insitute of Computer Science, FORTH,
N. Plastira 100, Vassilika Vouton,
GR70013, Heraklion, Crete, Greece

[2] Computer Science Department, University of Crete,
Heraklion, Crete, Greece

Figure 1: Method overview. Left: input depth and color frames. Middle: Object depth segmented using the fingertip 3D positions. Partially scanned object model and hand models. Right: 3D Rendering of the scene and final scanned model.

Experiment	Proposed mean/median error	[2], GT model mean/median error	[2], Scanned model mean/median error
Single hand, cat	0.42 / 0.39	0.47 / 0.43	0.45 / 0.43
Single hand, spray	0.65 / 0.63	0.70 / 0.53	0.63 / 0.47
Two hands, cat	0.38 / 0.34	0.33 / 0.31	0.44 / 0.39
Two hands, spray	0.59 / 0.44	0.51 / 0.38	0.62 / 0.41

Table 1: Hand tracking accuracy (in cm) measured on the synthetic datasets. The accuracy of the method is close to that of [2], although the latter assumes that the object model is known a priori.

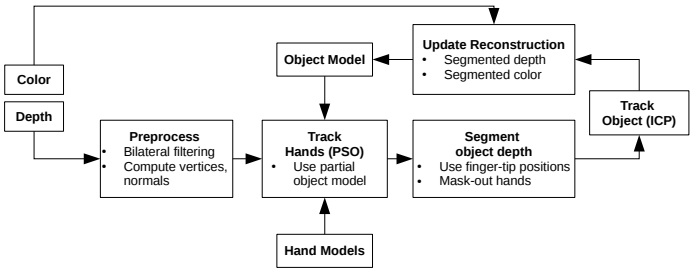

Figure 2: Work flow of the proposed method.

The analysis and the understanding of object manipulation scenarios based on computer vision techniques can be greatly facilitated if we can gain access to the full articulation of the manipulating hands and the 3D pose of the manipulated objects. Currently, there exist methods for tracking hands in interaction with objects whose 3D models are known [2]. There are also methods that can reconstruct 3D models of objects that are partially observable in each frame of a sequence [3]. However, no method can track hands in interaction with unknown objects, ie objects whose 3D model is not known a priori.

In this paper we propose a novel approach that can track human hands in interaction with unknown objects. As illustrated in Fig.1, the input to the method is a sequence of RGBD frames showing the interaction of one or two hands with an unknown object. Starting with the raw depth map (left) we perform a pre-processing step and compute the scene point cloud. We employ an appropriately modified model based hand tracker [4] and temporal information to track the hand 3D positions and posture (middle bottom). In this process, a progressively built object model is also taken into account to cope with hand-object occlusions. We use the estimated fingertip positions of the hand to segment the manipulated object from the rest of the scene (middle top). The segmented object points are used to update the object position and orientation in the current frame and are integrated into the object 3D representation (right).

More specifically, the work flow of the proposed approach consists of five main components linked together as shown in Fig. 2. At a first, preprocessing stage, the raw depth information from the sensor is prepared to enter the pipeline. A point cloud is computed along with the normals for each vertex. Then, the user's hands are tracked in the scene. An articulated model for the left and right hands, with 26 degrees of freedom each, is fit to the pre-processed depth input. The current, possibly incomplete (or even empty, for the first frame) object model is incorporated to hand tracking to assist in handling hand/object occlusions.

Using the computed 3D location of the user's hands as well as the last position of the (possibly incomplete) object model, the region of the object is segmented in the input depth map. The hands are masked-out from the observation, by comparing it to the rendered hand models.

Object tracking is achieved using a mutli-scale ICP [1]. The segmented object depth is used for a coarse to fine alignment with the (partially reconstructed) object model.

Finally, the segmented and aligned depth data of the object with the current, partial 3D model are merged. The object's 3D model is maintained in a voxel grid with a Truncated Signed Distance Function (TSDF) [3] representation.

After having processed all the frame of a given sequence, the 3D model of the object is reconstructed, provided that every part of the object was observed at at least one frame. Besides the accurate tracking of the hands, the proposed method provides an accurate 3D model of the object in the form of texture-mapped 3D mesh.

The proposed method was tested quantitatively and qualitatively in sequences where a person manipulates objects of different sizes, with either one or two hands. Table 1 shows the mean and median hand tracking error over a sequence with known ground truth. The first column shows that for the proposed method (object model is not known). The second shows that for [2] (perfectly accurate object model - ground truth - a priori known). The obtained results demonstrate that the hand tracking accuracy of our method is comparable to that o[2], although our method is not aware of the object model. Moreover, the comparison of the reconstructed object models to the actual ones shows only minor 3D reconstruction errors. Qualitative results obtained from a number of experiments are available at `http://youtu.be/9r43PtJ0Fwg`.

This work was partially supported by the EU FP7-ICT-288533 project ROBOHOW.COG.

[1] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[2] Nikolaos Kyriazis and Antonis Argyros. Scalable 3d tracking of multiple interacting objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2014.

[3] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[4] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis A. Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *BMVC*, Dundee, UK, Aug. 2011.