3D Object Reconstruction using Hand-Object Interactions

Rajat Vikram Singh
rajats@andrew.cmu.edu

Vaibhav Raheja
vaibhavr@andrew.cmu.edu

Carnegie Mellon University

In-Hand Scanning

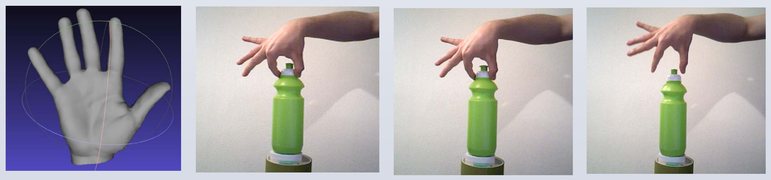
- **3D reconstruction possible using commercial RGB-D cameras:**
  - Moving Camera Reconstruction possible through KinectFusion which uses a rotating camera to reconstruct objects/spaces
  - Other approaches use this method for smaller objects with a static camera and
    - A turntable or
    - In-hand scanning
- **In-Hand Scanning:**
  - Uses high-temporal continuity
  - Needs prominent and stable texture information
  - Does not perform well in the absence of these features and
  - Rejects any information from the hands

Problem

- Objects which lack texture and geometrical information are challenging to reconstruct in 3-D
- Intrusive methods can be used here
- Hand-contact points approach works towards a non-intrusive method to reconstruct such objects

Available Data

- **Hand Model:**
  - Image frames where the hand interacts with the objects
  - Movement of hand encoded for all the frames using bone information and transformations for each frame
  - Structure of the hand, divided into 20 bones with a parent bone and length of bone information in a .SKEL file
  - 3D Mesh with vertices and polygons information in the form of .OFF file
  - Skin weights for each bone at each vertex in a .SKIN file



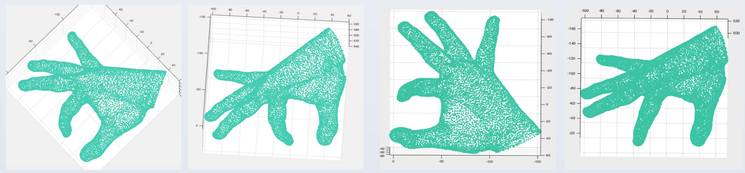
- **RGB-D Data**
  - RGB images which show the object being lifted by the hand
  - Depth information given in grayscale image format
  - Threshold images for easy visualization of depth

Approach

- Hand-Object Interactions based approach to extract contact points
- Find transformation of object using information from hand movement
- Fit the model of hand to data to get a pose
- Use the pose to find the contact correspondences which get discarded in traditional methods

Reconstruction

- **Hand Pose Estimation:**
  - Estimate the pose of the hand for each frame using transformations on the skeleton of the hand
  - Transform vertices of the mesh to get the pose of the hand in the current frame
  - Get a point-cloud of the hand



- **Object Point Cloud:**
  - Use the skin information to segment the hand out of the input images
  - Separate the object images using a mask of skin and depth threshold
  - Use the depth information to project the data into a 3D point-cloud

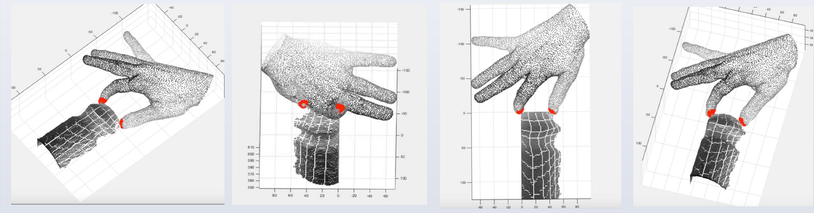
$$(X_{hand}, X'_{hand}) \in C_{hand}(\boldsymbol{\theta}, D_h, D_o, R, t) = E_{contact}(\boldsymbol{\theta}, D_h, D_o, R, t)$$

where X_{hand}, X'_{hand} are the corresponding contact points in source and target frame

- **Contact Points Estimation:**
  - Estimate points of contact of hand with the object
  - Find the distance between the object point cloud and the vertices of the hand point cloud
  - Find the vertices with distance less than 2 mm
  - Keep finding vertices until you find vertices for two fingers (found using the skin and skeleton information)

Reconstruction

- **Contact Point Estimation (cont.):**



- **Merging Point Clouds:**
  - Get touch correspondences from two consecutive frames
  - Find the transformation matrix and project the point cloud of the target frame using it
  - Merge object point clouds using Iterative Closest Pair (ICP) algorithm
  - The 3D point cloud is the reconstructed version of the object

Results

References

[1] 3-D Object Reconstruction from Hand-Object Reconstructions, Dmitri Tzionas, Juergan Gall.

[2] Hand Parsing for Fine-Grained Recognition of Human Grasps in Monocular Images, Akansha Saran, Damien Teney, Kris M. Katani.

[3] Motion capture of hands in action using discriminative salient points, L. Ballan, A. Taneja, J. Gall, L. Van Gool, M. Pollefeys.

[4] Tzionas, Dimitrios, et al. "Capturing hand motion with an RGB-D sensor, fusing a generative model with salient points." Pattern Recognition. Springer International Publishing, 2014. 277-289.