Completing 3D Object Shape from One Depth Image

Jason Rock, Tanmay Gupta, Justin Thorsen, JunYoung Gwak, Daeyun Shin, Derek Hoiem
Department of Computer Science, University of Illinois at Urbana-Champaign

Our goal is to recover a complete 3D model from a depth image of an object. Although we only see a small portion of an object from a single viewpoint, we can accurately predict the entire shape. This ability to infer complete 3D shape from a single view is important for grasping as we often reach around an object to grasp its unseen surfaces. Likewise, shape provides cues to category, affordance, and other properties. Recovery of 3D shape from a depth image is also useful for content creation and augmented reality applications. But how do we guess the shape of unseen surfaces? One approach is to recognize the same object or a similar object from past experience: the hidden surfaces of a favorite coffee mug can be inferred from earlier views or handling. Another is to infer missing surfaces using symmetries to duplicate and transform observed surfaces.

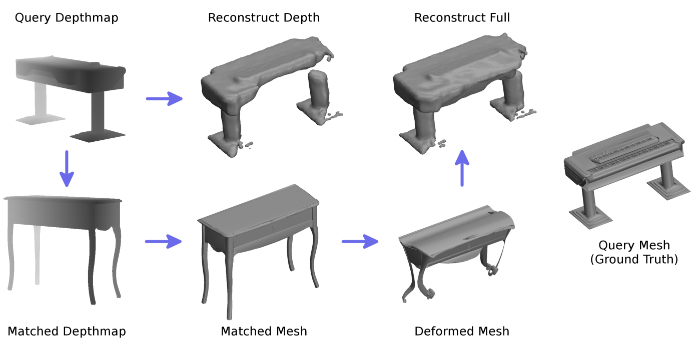

Figure 1: Pipeline for mesh reconstruction from depth image. A depth image of an piano is matched to a depth image of a table. The exemplar table is retrieved and deformed to better fit the observed depth points. Finally, a reconstructed mesh is created based on the observed depth points and deformed table exemplar. This mesh outperforms a mesh constructed from only depth points alone ("Reconstruct Depth").

Existing approaches rely on user interaction or apply to a limited class of objects, such as chairs. We aim to fully automatically reconstruct a 3D model from any category. We take an exemplar-based approach: retrieve similar objects in a database of 3D models using view-based matching and transfer the symmetries and surfaces from retrieved models. Given a depth image, our approach, illustrated in figure 1, is to first **retrieve** a similar depth image. The retrieved model is then **deformed** automatically to better approximate the depth image. Our final step, combines features from the depth image and shape priors from the exemplar mesh to **complete** an estimated mesh.

Retrieval: Our retrieval method is based on random forest hashing.[1] The random forest is trained to partition the training set into similar 3D shapes based on features of a depth image. Each tree of the forest acts as a hashing function, mapping the input features to a set of training examples. Since we wish to Random forest hashing provides sublinear retrieval from our large mesh-view dataset.

Deformation: We back project the query and retrieved depth maps into the world coordinates using the retrieved camera parameters. The two point clouds are aligned through a similarity transformation using pruned spin-image correspondences.[3] Plane symmetries are automatically detected in the retrieved mesh using a simplified version of [6]. We then deform the model using 3D approximation thin-plate splines (TPS) with additional symmetry constraints. The symmetry constraints allow us to deform unseen portions of many meshes.

Completion: We combine information directly observable from the depthmap, such as voxels we see through cannot be occupied, with cues from the matched mesh to produce a feature vector for a voxel map. We then predict the probability of occupancy using a boosted decision tree based on LogitBoost [2]. The probabilities are used as unary terms in a graph cut which encourages smoothness. Finally, we estimate a surface point-cloud from the voxel occupancy, and apply Poisson Reconstruction [4] to produce a visually pleasing result.

We investigate completion of 3D models in three cases: novel view (model in database); novel model (models for other objects of the same category in database); and novel category (no models from the category in database). We demonstrate reconstruction on a new synthetic dataset built from the SHREC12 mesh classification dataset.[5] The dataset consists of twenty meshes from sixty diverse classes, including musical instruments, buildings, and vehicles.

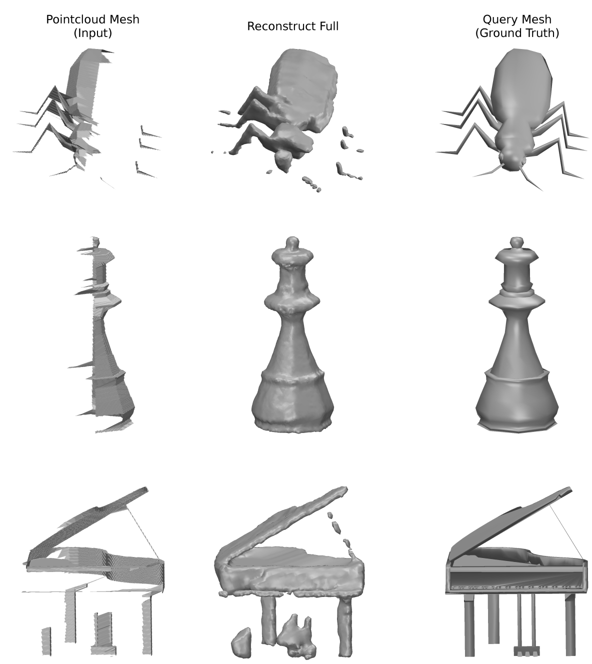

Figure 2: A selection of reconstructions: Much of the object is not directly visible from a single view ("Pointcloud Mesh"). Our result, which completes the object is shown in the second column, and the ground truth mesh is shown in the third.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Leo Breiman. Random forests. *Machine learning*, 2001.

[2] Michael Collins, Robert E Schapire, and Yoram Singer. Logistic Regression, AdaBoost and Bregman Distances. *Machine Learning*, 2002.

[3] Andrew E. Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *PAMI*, 1999.

[4] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Eurographics Symposium on Geometry Processing*, 2006.

[5] B. Li, Afzal Godil, Masaki Aono, X. Bai, Takahiko Furuya, L. Li, Roberto Javier López-Sastre, Henry Johan, Ryutarou Ohbuchi, Carolina Redondo-Cabrera, Atsushi Tatsuma, Tomohiro Yanagimachi, and S. Zhang. Shrec'12 track: Generic 3d shape retrieval. In *3DOR*, 2012.

[6] Niloy J Mitra, Leonidas J Guibas, and Mark Pauly. Partial and approximate symmetry detection for 3d geometry. *TOG*, 2006.