Supplemental Material: Co-Segmentation of Textured 3D Shapes with Sparse Annotations

Mehmet Ersin Yumer
Carnegie Mellon University
meyumer@cmu.edu

Won Chun
Google
wonchun@google.com

Ameesh Makadia
Google
makadia@google.com

1. Supplementary material

This document contains some details of our approach helpful for reproducing our method and results.

2. Implementation Details for Descriptors

2.1. Geometry Descriptors

Curvature. We compute principal surface curvatures k_1 and k_2 in the one ring neighborhood of a point[3]. Following [6] we construct a feature vector in $\mathbb{R}^9$ with the following values: k_1, $|k_1|$, k_2, $|k_2|$, $k_1 * k_2$, $|k_1 * k_2|$, $\frac{k_1+k_2}{2}$, $\frac{|k_1+k_2|}{2}$, $k_1 - k_2$.

Spin Image. The spin image is computed by parameterizing the local surface geometry around a point into the radial distance to the point's surface normal, and the signed distance to the point's tangent plane [5]. We compute a 6×6 histogram descriptor in this parameter space, with points weighted by their Voronoi area.

Shape Context. Originally used for shape matching in images [2], this descriptor computes a 2D histogram in relative log-geodesic distance and relative orientation of model points relative to the reference point. We use a 5×6 histogram with points weighted by their Voronoi area.

Neighborhood-PCA. Following [6] we compute the singular values σ_1, σ_2, σ_3, of the covariance matrix of local points at various geodesic radii ($5\%, 10\%$ and 25%). A 36 dimensional feature descriptor is then constructed from the following, computed in all 3 scales: $\frac{\sigma_1}{\sigma}$, $\frac{\sigma_2}{\sigma}$, $\frac{\sigma_3}{\sigma}$, $\frac{\sigma_1+\sigma_2}{\sigma}$, $\frac{\sigma_1+\sigma_3}{\sigma}$, $\frac{\sigma_2+\sigma_3}{\sigma}$, $\frac{\sigma_1}{\sigma_2}$, $\frac{\sigma_1}{\sigma_3}$, $\frac{\sigma_2}{\sigma_3}$, $\frac{\sigma_1}{\sigma_2} + \frac{\sigma_1}{\sigma_3}$, $\frac{\sigma_1}{\sigma_2} + \frac{\sigma_2}{\sigma_3}$, $\frac{\sigma_1}{\sigma_3} + \frac{\sigma_2}{\sigma_3}$, where $\sigma = \sigma_1 + \sigma_2 + \sigma_3$.

Average Geodesic Distance [4]. This descriptor indicates how isolated a point is from the rest of the model. In practice geodesic distances cannot always be computed (e.g. noisy models made up of multiple disconnected components). In such cases we follow [8] and introduce connections between components through contacts, followed by loosely adding edges between their closest points.

2.2. Appearance Descriptors

We incorporate the appearance cues implicit in the texture of the 3D models by processing the original texture image $\mathcal{I}$.

LAB Color. We convert the pixel colors into the perceptually uniform $L^*a^*b^*$ color space in order to use as features in Superpixelization (Section 3.3 in the paper).

Textons. We adopt the 17 filter bank and texton clustering of [7]. Specifically, we convolve the texture images of all models in the shape set with a bank of filters of size 5×5 which is composed of Gaussians with 3 different scales (1, 2, 4) applied to $L^*a^*b^*$ channels, Laplacian of Gaussians with 4 different scales (1, 2, 4, 8), and the derivatives of Gaussians with two different scales(2, 4) for each image axis. We then cluster all pixels of all texture images to generate *texton* descriptors (Figure 1(b) in the paper). Finally, we construct a texton histogram for each superpixel, from its pixels.

3. Generating sparse labels for untextured datasets

For datasets that do not have a sparse labeling input, we use multiple trials of randomly selected points. The process of how these points are selected is detailed below. If a segment S is made up of N faces that form a single connected component $S = \{f_i, i = 1, \ldots, N\}$, and $S_b \subseteq S$ denotes the boundary faces of the segment. If we let $d(f, S) = \min_{f_i \in S} d_g(f, f_i)$ (where d_g denotes geodesic distance between faces), then the centrality of a face $f_i \in S$ is given by $d(f_i, S_b)$. To select a face for a segment we can sample a point from a distribution, where the probability that a face is selected as the sparse representative is given by

$$p(f_i) = \frac{d(f_i, S_b)^{\frac{1}{\alpha}}}{\sum_i d(f_i, S_b)^{\frac{1}{\alpha}}}, \alpha \in (0, 1] \qquad (1)$$

When $\alpha = 1$, the probability of selecting face f_i as the segment representative is directly proportional to its distance to the segment boundary. As α approaches 0 the dis-

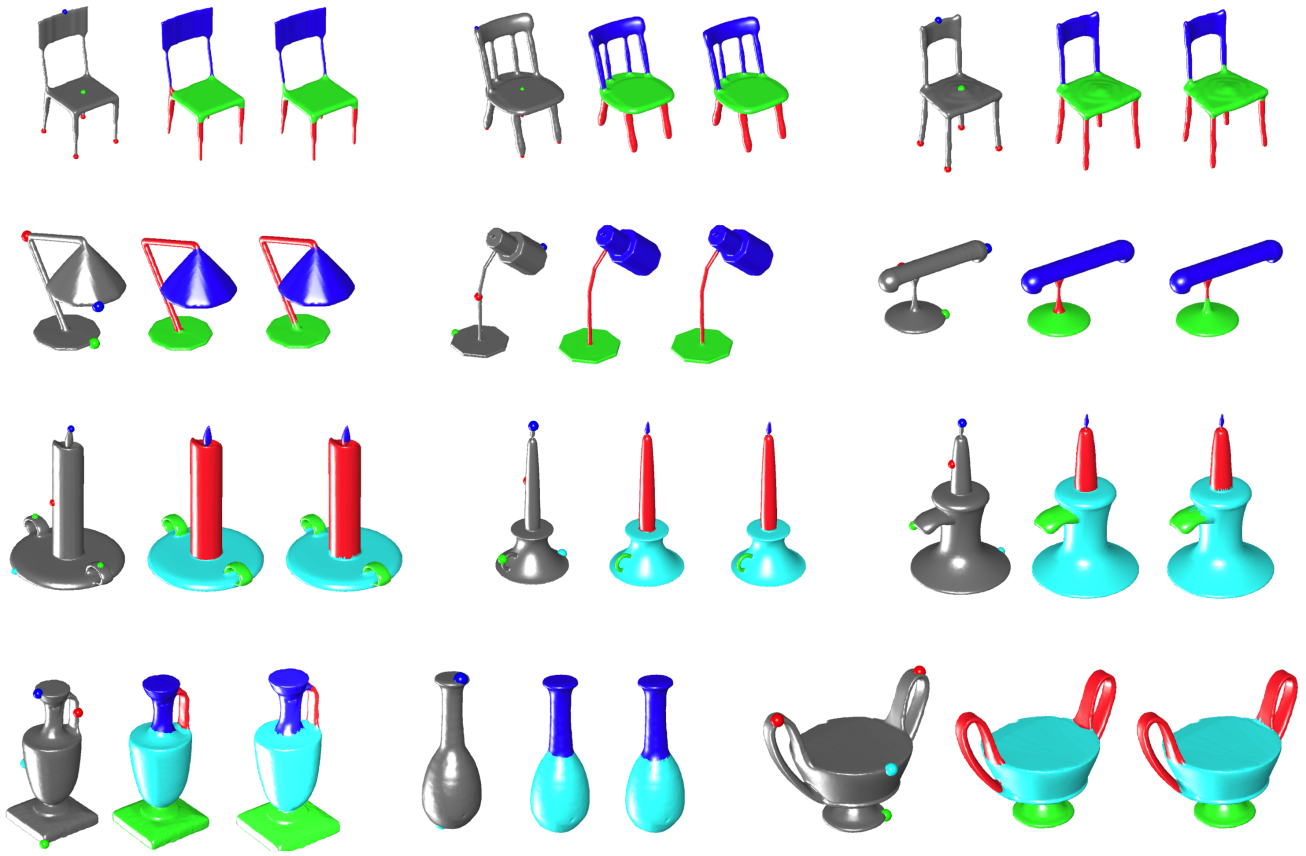

Figure 1. From top to bottom: Chairs, Lamps, Candelabra, Vases. From left to right: Input call-out points, ground truth, our segmentation.

tribution is further skewed so that we give even more preference to more central points (further from the boundary). In our experiments we use $\alpha = \{0.05, 0.25, 0.50, 0.75\}$.

4. Additional Results

4.1. Untextured data sets

Figure 1 illustrates results from each of the four categories used in the non-textured mesh segmentation experiments. The sparse labels are generated with $\alpha = 0.05$ using Equation 1.

5. Algorithm for 3D Superpixelization

The pseudo-code for superpixelization algorithm is given in Algorithm 1.

6. Foreground Segment Label Dictionaries of the Cameras and Video Recorders

Below, we list the part labels that are present in the cameras and video recorders datasets we used in the paper.

6.1. Cameras

Shutter Button , On/Off button , Zoom Button , Speaker , AF Assist Beam , Flash , Microphone , Battery Cover , Tripod Receptacle , LCD Screen , Mode Switch , Mode Dial , Exposure Compensation Button , Set Button , Display Button , Macro Button , Flash Button , Indicator , Playback Button , Menu Button , Terminal Cover , Strap Mount , Lens , Infrared Port , ISO Button , White Balance Button , Lens Release Button , Lens Mount , USB Port , Lens Contacts , Control Dial , Zoom Ring , Information Button , Delete Button , DC Coupler , View Finder , Hot Shoe.

6.2. Video Recorders

Microphone , Display , Lens , Speaker , HDMI Connector , USB Connector , Start/Stop Button , Status Indicator , Prev Button , Next Button , Lock Switch , Mic Terminal , Photo Button , On/Off button , Tripod Receptacle , Terminal Cover , Display Button , Playback Button , Play/Pause Button , BGM button , Auto Button , Memory Slot , A/V Out , Menu Button , Joystick , Lens Cover Release , Strap Mount , Access Lamp , DC Connector , Built-in Flash , A/V Out , Microphone jack.

Algorithm 1 3D Superpixelization

Detect active pixels in the texture as the pixels that are used in the appearance of the mesh (See Figure 2 in the paper for an example illustration).
for all Active pixels in the texture **do**
  Compute $L^*a^*b^*$ color: $[l_k, a_k, b_k]$.
  Compute 3D world position by projecting the pixel onto the triangle it appears on the mesh using the texture coordinates: $[x_k, y_k, z_k]$.
  Compute 3D world normal by projecting the pixel onto the triangle it appears on the mesh using the texture coordinates: $[n_k, n_k, n_k]$.
  Assemble feature vector using color, image coordinates, 3D world coordinates, and normals: $[l_k, a_k, b_k, x_k^I, y_k^I, x_k, y_k, z_k, n_k, n_k, n_k]$.
end for
Initialize cluster centers by sampling pixels at a regular grid step S for all pixels including inactive pixels: $C_k = [l_k, a_k, b_k, x_k^I, y_k^I, x_k, y_k, z_k, n_k, n_k, n_k]$.
Perturb cluster centers in an $n \times n$ neighborhood, to the lowest gradient position ($n \ll S$).
repeat
  for all Cluster centers **do**
    Assign best matching active pixels from a $2S \times 2S$ square neighborhood around the cluster center according to the distance measure (smallest distance with Equation 1 in the paper).
  end for
  Compute residual error E ($L1$ distance between previous centers and recomputed centers).
until $E \leq threshold$
Enforce connectivity (See [1] for details).

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. *École Polytechnique Fédéral de Lausssanne (EPFL), Tech. Rep*, 149300, 2010. 3

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, Apr. 2002. 1

[3] C.-S. Dong and G.-Z. Wang. Curvatures estimation on triangular mesh. *Journal of Zhejiang Uni. Science*, 2005. 1

[4] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii. Topology matching for fully automatic similarity estimation of 3d shapes. In *Computer Graphics and Interactive Techniques*, pages 203–212. ACM, 2001. 1

[5] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):433–449, 1999. 1

[6] E. Kalogerakis, A. Hertzmann, and K. Singh. Learning 3D Mesh Segmentation and Labeling. *ACM Transactions on Graphics*, 29(3), 2010. 1

[7] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. of Computer Vision*, 81(1), 2009. 1

[8] O. van Kaick, K. Xu, H. Zhang, Y. Wang, S. Sun, A. Shamir, and D. Cohen-Or. Co-hierarchical analysis of shape structures. *ACM Transactions on Graphics*, 32(4), 2013. 1