

Unsupervised Co-Segmentation of a Set of Shapes via Descriptor-Space Spectral Clustering

Oana Sidi* Oliver van Kaick[†] Yanir Kleiman* Hao Zhang[†] Daniel Cohen-Or*

*Tel-Aviv University

[†]Simon Fraser University

Abstract

We introduce an algorithm for unsupervised co-segmentation of a set of shapes so as to reveal the semantic shape parts and establish their correspondence across the set. The input set may exhibit significant shape variability where the shapes do not admit proper spatial alignment and the corresponding parts in any pair of shapes may be geometrically dissimilar. Our algorithm can handle such challenging input sets since, first, we perform co-analysis in a *descriptor space*, where a combination of shape descriptors relates the parts independently of their pose, location, and cardinality. Secondly, we exploit a key enabling feature of the input set, namely, dissimilar parts may be “linked” through third-parties present in the set. The links are derived from the pairwise similarities between the parts’ descriptors. To reveal such linkages, which may manifest themselves as anisotropic and non-linear structures in the descriptor space, we perform spectral clustering with the aid of diffusion maps. We show that with our approach, we are able to co-segment sets of shapes that possess significant variability, achieving results that are close to those of a supervised approach.

Keywords: Co-segmentation, shape correspondence, spectral clustering, diffusion maps.

Links: [DL](#) [PDF](#)

1 Introduction

In recent years, there has been an increasing interest in high-level analysis of 3D shapes. Some methods try to infer high-level knowledge about a given shape from its geometry [Fu et al. 2008; Mitra et al. 2010]. There are also works which utilize semantic knowledge to segment a given shape [Simari et al. 2009; Kalogerakis et al. 2010] or establish a correspondence between a pair of shapes [van Kaick et al. 2011]. The problem of analyzing a *set* of shapes as a whole has received less attention [Golovinskiy and Funkhouser 2009; Xu et al. 2010]. The interesting question about *co-analysis* of a set of shapes is whether more knowledge can be inferred from the set rather than from an individual or pairs of shapes alone. For example, can we better segment the shapes given as a set rather than as individuals? While it seems obvious that a set of shapes contains more knowledge than each individual, it remains a challenge, particularly in the *unsupervised* setting, to extract appropriate knowledge inherent to the set to facilitate fundamental analysis tasks such as segmentation and correspondence.



Figure 1: Unsupervised co-segmentation of a highly varied set of container objects using our algorithm. Corresponding parts differ significantly in their shape, pose, position, and cardinality.

In this paper, we investigate the problem of unsupervised co-segmentation of a set of shapes so as to reveal the semantic shape parts and establish their correspondence across the set. In our setting, the input shapes belong to a common family, that is, loosely speaking, they share the same functionality and general form. However, their corresponding parts are not necessarily similar; see Figure 1. The co-segmentation is unsupervised in that there is no training set which provides any knowledge to assist the analysis, as opposed to works such as [Kalogerakis et al. 2010] and [van Kaick et al. 2011], where the ability to properly match geometrically dissimilar parts is critically supported by the existence of relevant prior knowledge in the training set. It should also be noted that the analyses performed in the above works do not represent a co-analysis of a *set*. Our work can be seen to complement these knowledge-driven approaches with knowledge extracted from a target set.

The setting and objective of our analysis share similarities with the recent works of Golovinskiy and Funkhouser [2009] and Xu et al. [2010], which both compute an unsupervised co-segmentation of a set of shapes. The method in [Golovinskiy and Funkhouser 2009] pre-aligns the set of shapes and then combines criteria for intra-shape segmentation and inter-shape proximity to cluster mesh faces across the set. Xu et al. [2010] apply a similar co-segmentation scheme with the focus being to remove non-homogeneous part scales from the analysis equation. In our setting, the set consists of shapes with a variety of non-rigid, geometric, and even topological differences; see Figure 1. The dissimilarity between corresponding parts is reaching a point where the power of the set has to be exploited beyond what can be afforded by spatial-domain alignment or clustering. Towards this end, our approach allows correspondence to be inferred indirectly through *third parties* in the set and the analysis is performed in a *descriptor space*; see Figure 2.

Specifically, we treat the unsupervised co-segmentation of a set of shapes as a clustering problem. The clustering is performed in a space of shape descriptors rather than on the spatial coordinates of the shapes themselves. This allows the handling of corresponding

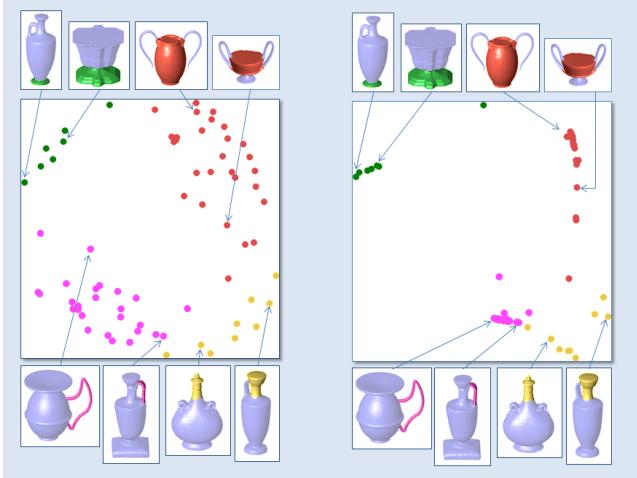


Figure 2: Challenge of unsupervised co-segmentation amid significant geometric variation (left) and effectiveness of descriptor-space spectral clustering (right). In the original descriptor space (left), two segments in the same semantic class (two pink handles) can be far apart, while unrelated segments (a pink handle and a yellow neck) can be closer. It is challenging to resolve this without any knowledge of the semantic classes. However, the handles are drawn close in the diffusion map (right) through third-party connections. The third parties, which are all the segments lying in-between the two handles, establish several paths between the two segments, given by the high similarities between pairs of points. These multiple paths create a strong connection between the two handles. Note that the two plots are 2D embeddings of the descriptor space, obtained with multidimensional scaling.

parts which may differ in pose, location, and even cardinality. Obviously, such variations would challenge any technique based on spatial alignment or direct clustering of shape geometry, e.g., [Golovinskiy and Funkhouser 2009], as shown in Figure 3. In addition, the descriptor clustering approach allows us to exploit a key enabling feature of the input set, namely, third-party connections. Even if two shapes possess parts that are significantly dissimilar, we can still establish a link between them if there are other parts in the set (third parties) that create such a connection, resulting in a successful co-segmentation. Figure 4 provides a concrete example. In contrast, spatial-domain alignment and clustering alone is unable to fully utilize the existence of third-party connections.

When performing the analysis in a descriptor space, the clusters that characterize the different shape parts do not necessarily take on an isotropic form; they may be elongated so that two corresponding parts which should belong to the same cluster are relatively far apart, e.g., see Figure 2. To this end, we perform spectral clustering with the aid of diffusion maps, which take the non-linear and anisotropic structures in the data and unfold them into a new space, so that the similarities between data points are translated into geometric proximity. Mean-shift clustering can then succeed in the embedded space by simply considering Euclidean distances.

Our algorithm starts by computing a per-object segmentation for each shape in the input set. For each segment on each shape, we compute a shape descriptor. Using the set of shape descriptors, all the obtained segments are then embedded into a common space via diffusion maps. By clustering the segments in this space, the initial co-segmentation is completed. Finally, the resulting clusters are used to create a statistical model to describe each class of parts, and a refined co-segmentation is obtained by labeling the shapes based

on the statistical models. This step allows us to correct imperfections that appear in the initial co-segmentation, e.g., at segmentation boundaries. Figure 5 illustrates the steps of our algorithm. We show that our unsupervised approach is able to co-segment families of shapes with significant geometric variations, achieving results that are competitive to supervised approaches [Kalogerakis et al. 2010; van Kaick et al. 2011]; e.g., see Figures 1 and 7.

2 Related work

One of the fronts in geometry processing that has been receiving much attention lately is high-level shape analysis, where the goal is to derive structural and semantic shape information from their low-level geometric properties. For example, some methods extract a set of representative primitives from a shape, such as structural curves [Gal et al. 2009] or rigid components and joints [Xu et al. 2009], so that the shape’s defining characteristics are preserved during deformation. Other achievements in high-level analysis include finding the *upright orientation* of shapes [Fu et al. 2008], a property closely related to the functionality of the objects, automatically analyzing an assembly of mechanical parts to infer their motion and chain of interaction [Mitra et al. 2010], and building a symmetry-induced hierarchy of man-made shapes [Wang et al. 2011].

Shape segmentation [Shamir 2008] and correspondence [van Kaick et al. 2010] are two of the most fundamental problems in high-level shape analysis. To segment a shape into meaningful parts, and to compare or match these parts, we need to understand the high-level structure of the shapes. Much effort has been devoted to solving these problems, however, the endeavor has been mainly focused on analyzing only one shape or a pair of shapes at a time.

The question that naturally follows is whether we can benefit from simultaneously analyzing a set of shapes from the same family, since intuitively more information is then available. In the im-

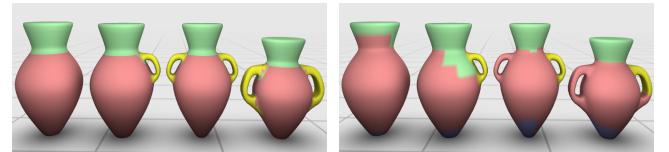


Figure 3: Co-segmentation in descriptor space (left) vs. in spatial domain [Golovinskiy and Funkhouser 2009] (right). The use of descriptor-space clustering enables our method to handle variations in part placement and cardinality (see the handles). Co-segmentation results via spatial alignment are less meaningful.

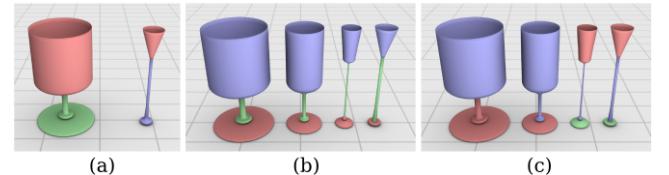


Figure 4: The power of third-party connections. When presented with a set containing only two rather dissimilar shapes (a), our co-segmentation scheme returns a less-than-satisfactory result. However, when the set is augmented with the two “in-between” shapes (b), they serve as third-parties which help establish a link between the dissimilar parts, providing a meaningful co-segmentation of the set. In contrast, a method based on spatial alignment and clustering [Golovinskiy and Funkhouser 2009] does not benefit from this property and leads to a less meaningful co-segmentation (c).

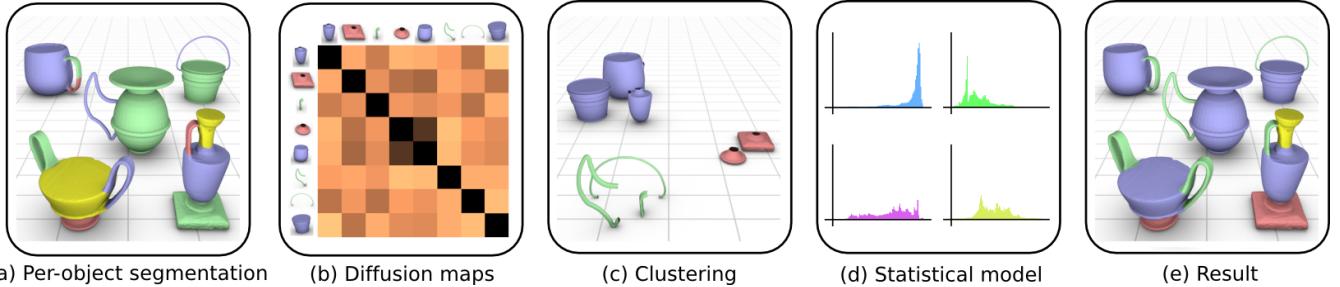


Figure 5: Overview of the steps in our co-analysis: (a) An individual segmentation is computed for each shape. (b) The segments from all the shapes are embedded into a common space by using diffusion maps based on a similarity matrix, where darker colors in the matrix entries indicate higher similarities. (c) The segments are clustered in the embedded space. (d) A statistical model is created to describe each cluster. (e) The statistical model is used to label the shapes and obtain the final co-segmentation of the set.

age domain, positive results for co-segmentation have been demonstrated, e.g., by using a generative model to match the appearance histogram of two images and enforce spatial coherency [Rother et al. 2006], or by making use of discriminative clustering, which seeks maximal separation of the classes [Joulin et al. 2010].

In the case of shapes, a first indication of the advantage of set analysis appears in the works of Kalogerakis et al. [2010] and van Kaick et al. [2011], where the knowledge is represented in the form of discriminative part models which are learned from a set of manually segmented and labeled shapes. The models can then be used to segment and label an unknown shape from the same class [Kalogerakis et al. 2010], or to establish a part correspondence between a pair of shapes [van Kaick et al. 2011]. Although the knowledge is defined in terms of a set of shapes, it is still manually created and not automatically inferred from the set.

The works of Golovinskiy et al. [2009] and Xu et al. [2010] take a concrete step towards co-analysis of shapes and propose methods that infer knowledge from the set alone. In [Golovinskiy and Funkhouser 2009], the co-segmentation is posed as a graph clustering problem. In this graph, each node corresponds to a face in one of the meshes, and the edges come from: (1) the individual connectivity of the shapes, and (2) from a set of correspondence edges connecting faces that are geometrically close. The correspondence edges are added after the shapes were aligned to each other. The clustering of this graph naturally provides a per-shape segmentation that is coherent across the group; the method is however limited to shapes that can be spatially aligned; see Figures 3 and 4.

To overcome limitations of global alignment and scaling [Kazhdan et al. 2004; Golovinskiy and Funkhouser 2009], Xu et al. [2010] classify the input shapes into different *styles* according to the scales of the shape parts. Co-segmentation based on graph clustering is then applied only to the shapes within each style cluster, and is modified to take into consideration the parts and derive the correspondence edges from part similarity. This modification allows the co-analysis to succeed for a larger variety of shapes, especially those whose parts differ by non-homogeneous part scaling.

We are interested in co-segmenting shapes with more variability than those in [Golovinskiy and Funkhouser 2009; Xu et al. 2010], such that corresponding parts can be rather dissimilar geometrically as well as topologically. Therefore, we deviate from the scheme of spatial alignment or clustering the shape parts in the spatial domain where the shapes reside. Instead, we automatically derive statistical models that describe the different parts of the shapes in a space of shape descriptors and utilize spectral clustering to account for clusters of arbitrary shapes.

In an independent work published in this issue, Huang et al. [2011]

describe an unsupervised algorithm that jointly segments shapes in a heterogeneous shape library, obtaining results comparable to supervised approaches on a benchmark test. However, unlike our approach, this technique does not guarantee that the segmentations of all shapes within a shape class are consistent.

There has been a large body of works in geometry processing using spectral methods [Zhang et al. 2010]. Spectral clustering via diffusion maps is also not new. The mesh segmentation work of de Goes et al. [2008] specifically applies the diffusion distance, as we do in our work. All these solutions exploit special properties of the spectral embedding. However, the computed embeddings have always been a transformation from the spatial coordinates of individual input shapes. Our work computes spectral embeddings of shape descriptors. The descriptors of all the shapes in the set take part in the embedding, enabling us to perform a co-analysis.

3 Overview

Our co-analysis method takes as input a set of meshes from a given family and computes their co-segmentation and labeling. The co-analysis also provides a correspondence among the segments of any pair or group of shapes in the set, since the segments corresponding to the same part class will possess a common label. The label does not necessarily carry a semantic meaning, but serves more as a part index. For each family of shapes, the user also provides the *maximum* number of labels L that should be recovered from the set. This number loosely corresponds to the number of different kinds of semantic parts that constitute the shapes, e.g., $L = 4$, for a set of vases that can have a base, body, handle, and neck. Note that some types of parts can repeat or be omitted on the shapes, e.g., there can exist vases with multiple handles and vases without a base.

To carry out the co-analysis, we start with a per-object segmentation of each shape in the input set. Then we extract shape descriptors for the initial sets of segments. Next, based on the descriptors, we cluster the segments using diffusion maps. Finally, we build a statistical model for each cluster, which is used to obtain the final co-segmentation and labeling of the shapes in the set. We describe these steps as follows (see Figure 5 for an illustration).

Per-object segmentation. The first step in the co-analysis is to obtain an individual segmentation for each shape in the input set. We achieve this by grouping the mesh faces with mean-shift clustering based on shape descriptors defined for the faces, although any reasonable alternative can be used here. More details are given in Section 4. The outcome of this procedure is a set of candidate segments per shape. The purpose of the per-object segmentation is to facilitate the co-analysis, since we are interested in analyzing shape

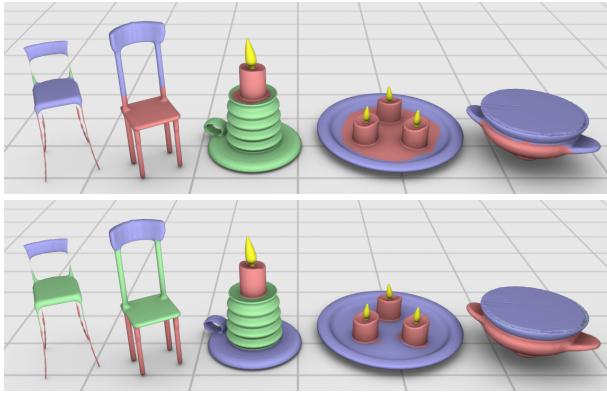


Figure 6: Result of the co-segmentation refinement (bottom) applied on an initial co-segmentation (top). Notice how boundaries are displaced to better locations (candles) and some mislabeled segments receive the correct label after the refinement (chairs). The co-segmentation is performed independently for each class.

parts, rather than lower-level primitives. However, notice that the final co-segmentation given by our algorithm is a refined labeling performed at the face level (Section 6), which allows us to correct imperfections that appear in the per-object segmentation.

Descriptor-space spectral clustering. All the candidate segments produced by the per-object segmentation are embedded into a common space via diffusion maps. The embedding translates similarity between segments into spatial proximity, so that a clustering method based on Euclidean distance will be able to find meaningful clusters in this space; see Figure 2 for an example. To define the embedding, an affinity matrix is constructed from the similarities between pairs of segments. The similarities are given by the distance between descriptors defined at the segment level. The embedding is then given by the first few scaled eigenvectors of the affinity matrix. Note that the embedding is constructed from the similarities among *all* the segments in the set. This implies that, even if two corresponding parts are far apart in the original descriptor space, they may end up close-by in the embedding if other parts in the set (third parties) imply their correspondence by transitivity. We obtain an initial co-segmentation by clustering the segments in the embedded space. All the segments in a single cluster potentially represent a certain class of parts, e.g., base, body, handle, or neck of a vase. This step is carried out with a hierarchical clustering scheme. We elaborate on descriptor-space clustering in Section 5.

Statistical model and refined co-segmentation. A statistical model is then built to describe each cluster of parts and used to label the shapes. The advantage of this final step of the method is that it allows us to correct eventual errors that appear in the initial segmentation, since now we perform a more detailed labeling of the shapes (Figure 6). More specifically, the statistical model is defined in terms of Gaussians learned from the shape descriptors. We derive the probability of labeling each face with a given class by how well the descriptors of the face fit the part model. Next, we obtain the best segmentation for each shape by applying graph cuts labeling based on these probabilities. The final result is a refined co-segmentation and coherent labeling of all the shapes in the set. More details on this step of the algorithm are given in Section 6.

4 Descriptors and per-object segmentation

Shape descriptor extraction. We normalize the shapes for overall scale, and compute a set of shape descriptors that are used in the dif-

ferent steps of the co-analysis. The algorithm works with faces and segments (a group of connected faces). Thus, we compute descriptors both at the face-level and at the segment-level. First, we obtain an upright orientation of each shape [Fu et al. 2008] and define a subset of the face-level descriptors based on the reoriented shapes. More specifically, we record for each face, the geodesic distance from the base of the shape to the face, and the angle between the normal of the face and the upright orientation vector. Finally, we also make use of the shape diameter function [Shapira et al. 2009], which gives an estimate of the thickness of the shape at the face.

For the segment-level descriptors, we compute histograms that capture the distribution of each face-level descriptor for all the faces in the segment. For a segment s_i and the face-level descriptor d , we denote the histogram as h_i^d . We also include two descriptors only defined at the segment level. a_i is the segment area normalized by the total shape area. g_i is a vector of three components that describes the overall geometry of the segment. We have $g_i = [\mu_l \mu_p \mu_s]$, where

$$\mu_l = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \mu_p = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{and} \quad (1)$$

$$\mu_s = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \quad \text{with } \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0. \quad (2)$$

These terms give an indication of how linear (cigar-shaped), planar and spherical the shape of the segment is. λ_1 , λ_2 , and λ_3 are the three eigenvalues obtained when applying principal component analysis to all the vertices that are part of the segment.

Per-object segmentation. Using the shape descriptors, we compute a per-object segmentation for each shape in the set. Although any reasonable segmentation algorithm can be used for this step, we opted to use the *mean-shift algorithm* [Comaniciu and Meer 2002], where we cluster the mesh faces into larger segments. Mean-shift operates by finding the modes (local maxima of density or cluster centers) of points in feature space. The advantage of using mean-shift lies in its non-parametric nature, i.e., the number of clusters does not have to be known in advance. Instead, the algorithm requires an estimation of the *bandwidth* or radius of support around the neighborhood of a point. However, this can be typically estimated from the data, e.g., as done in shape analysis [Shamir et al. 2006]. Following a similar procedure, we fix the bandwidth in our experiments to a percentage of the range of each descriptor. The actual distance measure used for the clustering is derived from the descriptors, but defined in terms of diffusion distances (Section 5). Finally, disconnected clusters are broken into separate segments according to their connectivity.

The output of this step is a set of candidate segments for each object. Notice that the result can constitute an over-segmentation of the shapes, i.e., a semantic part might be composed of more than one segment. However, our goal here is to obtain candidate segments as suggested by the shape descriptors. These segments will be utilized in the subsequent co-analysis of the set, which will then provide a refined segmentation and coherent labeling of the shapes.

5 Descriptor-space spectral clustering

We recall that one of our goals in the co-analysis is to extract information on what types of parts compose the shapes in the set. With this objective in mind, we take the segments computed individually for each shape (as described in the previous section) and cluster them into groups of similar segments. Next, we derive a statistical model for each cluster to represent each type of part that appears in the set. More details are given in Section 6.

However, it is important to notice that the objects that we consider can have a large amount of variability. Thus, a simple clustering algorithm will not group the segments into the proper classes, since the segments are non-uniformly distributed in descriptor space. The shape of the clusters can be highly anisotropic or even non-linear. To overcome this difficulty, we first embed the segments into a new space with the aid of diffusion maps, where the Euclidean distance between two segments will better reflect their similarity. Then, clustering in the embedding will provide a more accurate grouping of segments. The usefulness of applying the diffusion maps for the co-segmentation of a group of shapes is illustrated in Figure 2.

Embedding computation. We start with the set of segments obtained from all the shapes, $S = \{s_1, \dots, s_n\}$. The dissimilarity between two segments s_i and s_j is given by

$$\mathcal{D}(s_i, s_j) = \sqrt{\sum_{d=1}^{n_d} \text{EMD}^2(h_i^d, h_j^d) + |a_i - a_j|^2 + \|g_i - g_j\|_2^2}, \quad (3)$$

where h_i^d , a_i , and g_i are the segment-level descriptors discussed in Section 4 and $n_d = 3$. EMD is the earth-mover's distance, a common measure of the dissimilarity between two probability distributions, since the h_i^d are histograms.

Next, we construct an affinity matrix W , with

$$W_{i,j} = \exp(-\mathcal{D}(s_i, s_j)/2\sigma^2). \quad (4)$$

Note that, in our method, we obtain the pairwise affinities by applying a Gaussian kernel to the segment dissimilarities, but other choices for the kernel are also possible [Nadler et al. 2005]. By defining a matrix $D_{i,i} = \sum_j W_{i,j}$, we obtain the normalized $M = D^{-1}W$. M can be seen as a stochastic matrix, with $M_{i,j}$ being the probability of a transition from segment s_i to segment s_j in one time step. The transition probability can be interpreted as the strength of the connection between the two segments.

Finally, we compute the eigendecomposition of the matrix M , obtaining eigenvalues $\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n-1} \geq 0$ and eigenvectors $\psi_0, \dots, \psi_{n-1}$ [Nadler et al. 2005; Coifman and Lafon 2006]. The diffusion map at time t is then given by

$$\Psi_t(s) = (\lambda_1^t \psi_1(s), \dots, \lambda_{n-1}^t \psi_{n-1}(s)), \quad (5)$$

where $\Psi_t(s)$ defines the coordinates of segment s in the embedding or map. The eigenvector ψ_0 is constant and is thus discarded.

Interpretation. The main result regarding diffusion maps is that the Euclidean distance between two points x and y on the map is equal to the *diffusion distance* between the two points [Nadler et al. 2005]. The diffusion distance is given by

$$\mathcal{D}_t^2(x, y) = \sum_z (p(t, z|x) - p(t, z|y))^2 w(z), \quad (6)$$

where $p(t, z|x)$ is the probability of transition from x to z in t time steps, and $w(z)$ weights the local density at z . The larger the number of short paths that exist between x and y , the more the distance will decrease, giving an indication of how strongly the two points are connected. The time parameter t can be varied to analyze the structure of the points at different scales [Coifman and Lafon 2006]. An alternative interpretation is to see the map as the state of a dynamic system after t steps of a diffusion process have taken place.

Implementation. We compute the diffusion map using only the first three eigenvectors, since the diffusion distance can be well approximated in this manner [Nadler et al. 2005]. Moreover, we select

$t = 3$ for computing the embedding of all the segments and $t = 5$ for the initial per-object segmentation (as described in Section 4).

Clustering. After obtaining the diffusion maps, we cluster the segments in the embedded space with an agglomerative hierarchical algorithm. We start with each segment as an initial cluster and, during the incremental construction of the hierarchy, we merge the current pair of clusters with minimal distance. The distance between two clusters is given by the Euclidean distance between their centroids. The final number of clusters is provided by the user, and it corresponds approximately to the number of semantic parts that constitute the shapes. The result is a grouping of the segments into the potential classes of parts that exist in the set.

6 Statistical model and co-segmentation

Statistical model. We now derive a statistical model for each class of parts. The models are constructed from the clusters of segments, based on the shape descriptors.

For each cluster c_i , we collect the descriptor values for all the faces of all the segments in the cluster. Based on the observed values, we estimate a multi-dimensional Gaussian to model the class,

$$p(f|c_i) = p(f, \mu_i, \Sigma_i) = C e^{-\frac{1}{2}(f-\mu_i)^T \Sigma_i^{-1} (f-\mu_i)}, \quad (7)$$

where μ_i and Σ_i are the parameters that model the i -th class, and C is the Gaussian's normalization constant. The parameter dimensions are 3×1 and 3×3 , respectively, as we use three face-level descriptors. The parameters are estimated with a standard expectation-maximization approach. We observed that this scheme provides a simple, yet effective, class model, e.g., as opposed to more complex models such as mixtures of Gaussians.

Finally, the probability that an unknown face f belongs to class c_i is given by Bayes' Theorem,

$$p(c_i|f) \propto p(f|c_i)p(c_i), \quad (8)$$

where the prior $p(c_i)$ is taken as the sum of the area of the segments that are part of cluster c_i , normalized by the total area of all segments in the set.

Refined co-segmentation. The statistical model for each class is used to perform the final co-segmentation. We pose the co-segmentation as a labeling optimization which is solved individually for each shape. The group information enters the optimization through the data term of the labeling energy.

Given a mesh, we define the graph $G = \{V, E\}$, where the nodes V are given by the faces of the mesh and an arc $\{u, v\} \in E$ if the faces u and v are neighbors on the mesh. The optimization is then posed as finding the labeling l that minimizes the energy

$$\mathcal{E}(l) = \sum_{u \in V} \mathcal{E}_D(u, l_u) + \sum_{uv \in E} \mathcal{E}_S(u, v, l_u, l_v), \quad (9)$$

where l_u and l_v are the labels assigned to nodes u and v , respectively, and \mathcal{E}_D and \mathcal{E}_S are the data and smoothness energy terms.

The data term is given by

$$\mathcal{E}_D(u, l_u) = -\omega \log(p(c_{l_u}|u)), \quad (10)$$

where $p(c_{l_u}|u)$ is the probability that node u is part of cluster c_{l_u} , given by the statistical model of cluster c_{l_u} , and ω is a constant that regulates the influence of the data term in the total energy. The cost of assigning a specific label to the node increases according to how unlikely it is that the face belongs to the corresponding class.

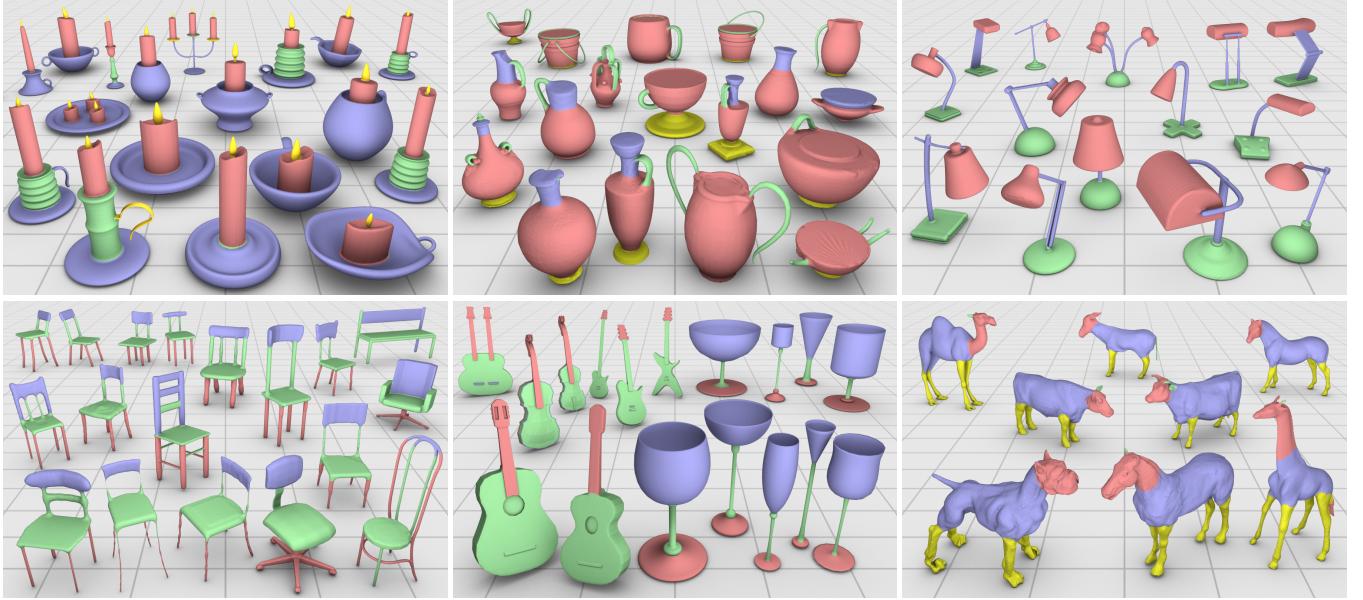


Figure 7: Results of our co-segmentation on a variety of shapes. Corresponding segments in each class are shown with the same color. Notice how the segmentation and labeling is coherent for many of the parts in each set. The results for all the sets were obtained with the same parameters and shape descriptors. Labeling accuracy statistics for each class are shown in Table 1.

Similarly to [Shapira et al. 2009], the smoothness term is defined as

$$\mathcal{E}_S(u, v, l_u, l_v) = \begin{cases} 0, & \text{if } l_u = l_v \\ -\log(\theta_{uv}/\pi) l_{uv}, & \text{otherwise,} \end{cases} \quad (11)$$

where l_{uv} is the length of the edge between the faces corresponding to u and v , and θ_{uv} is the dihedral angle between the two faces.

To obtain the labeling that minimizes the energy \mathcal{E} , we use graph cuts optimization [Boykov et al. 2001]. More specifically, the multi-label α -expansion algorithm is utilized.

The result of the labeling optimization is a co-segmentation of the set, since the labeling energy is based on the statistical models obtained from the co-analysis. Thus, in addition to an individual segmentation for each shape, we also obtain a correspondence among the segments, so that two segments on two different shapes correspond to each other if they possess the same label. Notice also that this step allows us to correct errors that appear in the initial co-segmentation, such as displaced boundaries and mislabeled segments. An example of the refinement is shown in Figure 6.

7 Results

In this section, we evaluate our unsupervised co-segmentation method and present qualitative and quantitative results. We also compare our method with the state-of-the-art in co-segmentation and with a supervised approach. Since our goal in this work is to present a fully unsupervised approach, all the results were obtained with a fixed set of parameters.

Datasets and methodology. We use seven classes of shapes in our experiments: candelabra, chairs, four-legged animals, goblets, guitars, lamps, and vases. The sets of man-made shapes are composed of objects that possess significant variability, i.e., a common type of part can appear with different topologies and geometries across the set, and it can be absent or appear multiple times on a shape. We modeled two new classes of man-made shapes for this work, while

the remaining classes appeared in [van Kaick et al. 2011]. We manually segmented and labeled each shape, according to a specific labeling scheme for each class. This provides a ground-truth label for each face. We also selected one set of organic shapes (four-legged animals) from the Princeton Segmentation Benchmark [Chen et al. 2009], and use the ground-truth labeling created by Kalogerakis et al. [2010]. Note that the ground truth is only used for a statistical evaluation and is not utilized by our algorithm. The co-segmentation is performed separately for each class.

Co-segmentation and labeling. Visual results of our co-segmentation are shown in Figure 7. Notice how, despite the great variability in the shape parts, our co-analysis is able to extract the common parts in the set and yield a coherent labeling. We point out illustrative examples. For candelabra, the method is able to identify the flames and wax candles of different sizes across the set, and separate them from the bases and holders, which appear in rich varieties. Also, the multiplicity of the flames and candles does not pose a problem to the method. Moreover, the lamps and goblets are successfully segmented and labeled into their three constituent parts, even though the geometry or thickness of the corresponding parts varies. Notice also the different topologies of the lamp supports that are detected. Chairs are also co-segmented into their main constituent parts, and we obtain a correspondence between legs of different topologies, including star-shaped legs and legs with railings. In the vases, handles of different sizes and shapes are identified, as well as bodies with very different geometries, including spherical, cylindrical, or flat bodies. We also see how the various types of guitar bodies are separated from the necks. Finally, we observe that the co-analysis is also successful when applied to a set of organic shapes with significant variability. Heads with short or long necks are correctly labeled, as well as the various animal bodies and legs.

We also notice a few shortcomings in the results. The small handles that appear on the candelabra are not properly separated from the holders, and similarly the small guitar headstocks are not separated from the fretboards. The cylindrical sections in the chair backrests are assigned to the same clusters as the seats. And, the animal tails are fused with the segments that represent the bodies, while some of

the ears and horns are fused with the heads and necks. These problems appear due to imperfections in the clustering, which assigns these parts with an incorrect label. Another shortcoming includes the chair made up of thin wires on the front row, which is mislabeled. In this case, we attribute the problem to the lack of information available to the co-segmentation: this shape has unique parts that do not have similar counterparts in the set. This is an intrinsic limitation of our approach.

Moreover, to assess the quality of the results in a quantitative manner, we show a statistical evaluation in Table 1. The first column indicates the number of shapes in the class, the second column corresponds to the labeling accuracy of the initial co-segmentation, and the third column shows the labeling accuracy after the refined labeling. Each entry is the average accuracy for all the shapes in the class. The accuracy for a single shape is given by

$$\text{Accuracy}(l, t) = \frac{\sum_i a_i \delta(l_i = t_i)}{\sum_i a_i}, \quad (12)$$

where a_i is the area of face i , l is the labeling returned by the co-segmentation, t is the ground-truth labeling, and $\delta(x = y)$ is 1 only if $x = y$. This measure captures the amount of area of the shape that is labeled correctly by the co-segmentation [Kalogerakis et al. 2010]. Since our co-segmentation does not return labels associated to specific semantic classes, before computing the accuracy we find the best one-to-one matching between our labels and the ground-truth labels. The matching is used coherently for the whole set.

The average labeling accuracy for all the classes is about 84% for the initial co-segmentation, and 88% for the refined labeling. We notice that, with the exception of one class, the accuracies are at least 84% or higher. We attribute the 10% accuracy gap between certain classes, e.g., candelabra vs. lamps, to the greater part variability that appears on the candelabra, chairs, and vases. Notice also a 4% improvement from the initial to the refined co-segmentation. This difference arises as a result of the boundary refinement, where parts that were oversegmented or wrongly labeled in the initial co-segmentation are properly labeled and have their segmentation refined with the statistical models (Figure 6).

Effect of the set. In Figure 8, we demonstrate the power of the set by evaluating how the co-segmentation accuracy improves when the set is enriched. For each class, we start with a pair of shapes and incrementally increase the set size by adding one shape at a time to the set. To isolate the effect of the graph cuts refinement, we perform only the initial labeling for each subset and report the labeling accuracy according to (12). The x -axis denotes the subset size, while the y -axis is the accuracy for processing the subset.

For all the classes, the general trend of the curves clearly demonstrates that the co-segmentation accuracy is improved as more shapes are added to the set. The non-monotonicity of the accuracy plots occurs when a difficult or unique exemplar is added to the set. The accuracy is then recovered or improved once additional shapes providing more information are added to the set. Note that the curves can change according to the specific order in which the shapes are added to the set. We particularly chose orderings that better reveal the monotonic behavior of the algorithm. Random orderings lead to more points that break the monotonicity, however, the overall upward trend of the curves remains the same.

Comparison to the state-of-the-art. Figure 9 shows the results obtained by applying the method of Golovinskiy and Funkhouser [2009] on two selected sets¹. Their results on all the other datasets are available in the supplementary material. Notice that this method obtains satisfactory results for sets that can be

Table 1: Average co-analysis labeling accuracy.

Class	Num. shapes	Initial lab.	Refined lab.
Candelabra	28	73.0	84.4
Chairs	20	78.6	84.8
Four-legged	20	75.9	77.3
Goblets	12	98.0	98.2
Guitars	44	86.4	87.2
Lamps	20	93.8	94.3
Vases	28	84.5	87.4

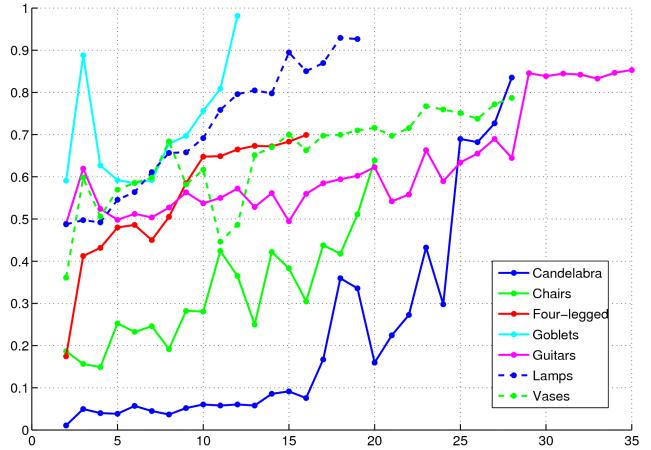


Figure 8: The power of the set demonstrated for all the classes: as the number of shapes in the set increases (x -axis), the accuracy of the co-segmentation (y -axis) shows clear trend of improvement.

properly aligned with similarity transformations, e.g., chairs. However, when the shapes possess significant variability in topology (vases and candelabra) or pose (four-legged animals), the segmentations either miss important parts of the shapes (the animal heads), or are not meaningful at all (as in the vases and candelabra). This is due to the fact that this method derives the relation between the parts of different shapes from their proximity after alignment. It is not possible, even after a perfect alignment, to derive a correspondence between the parts of some of the vases just from proximity. Our method, on the other hand, is more robust in this regard, since it makes use of shape descriptors to handle more shape variability, and derives the inter-shape relations from third parties in the set.

Comparison to a supervised approach. We also compare our unsupervised co-segmentation to the supervised approach of Kalogerakis et al. [2010]. The authors of this approach kindly provided the results of applying their method on four sets that we use in this work. In each experiment, 70% of the shapes in a set were randomly selected as training data, while the remaining 30% were used as test shapes on which the labeling accuracy was evaluated. These experiments were repeated 5 times for each class, and the results were averaged. The accuracy for a single shape was also computed with the measure in (12). The accuracies obtained are: 80.9%, 91.6%, 97.3%, and 96.0%, for candelabra, chairs, lamps, and vases, respectively. Although a direct comparison of the accuracies of both approaches does not represent a meticulous evaluation, due to the splitting of the dataset into train and test sets with the supervised approach, we believe that in this manner we are nevertheless capturing the accuracy of the supervised approach in an average case.

By comparing these numbers with Table 1, we see a difference of at most 10% between the two approaches. We conclude that such

¹The authors kindly provided their implementation to us.

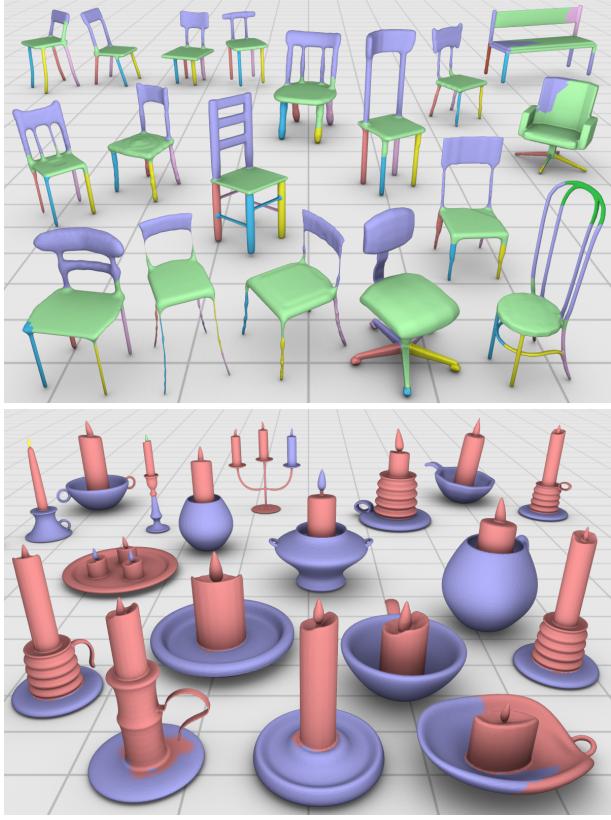


Figure 9: Comparison to the approach in [Golovinskiy and Funkhouser 2009]. In contrast to our co-analysis results in Figure 7, we observe that this method provides good results for sets of shapes that can be spatially aligned (chairs), but less meaningful co-segmentations for shapes with great part variability (candles).

close results demonstrate the high potential of the unsupervised co-segmentation. Although the results of the supervised approach are more accurate on average, we recall that it requires substantial work for the preparation of a reasonably-sized training set, since the shapes have to be manually segmented and labeled. Additionally, a separate training set is necessary for each different class of shapes, and the approach only extracts a labeling that follows the pattern of the training examples. If a different training set is given, the results can change considerably. The unsupervised approach, on the other hand, derives its knowledge automatically from the set. Finally, we conjecture that the accuracies for the unsupervised approach can possibly be increased with the enhancement of certain components of the method, such as the shape descriptors and statistical models.

Performance. Our implementation is fairly efficient, executing in 10 minutes for a set of 30 shapes in an AMD Opteron 2.4GHz with 8Gb of memory. In contrast, supervised approaches such as those in [Kalogerakis et al. 2010] and [van Kaick et al. 2011] can take on the order of hours for similarly-sized training sets and hardware configuration. Our faster performance is mainly due to two aspects: 1. The supervised approaches involve classifier training with complex algorithms such as *boosting*. 2. We perform most of the analysis at the segment-level, rather than face-level, a natural approach to greatly reduce the complexity of the method.

8 Conclusion, discussion, and future work

We presented a method for co-segmentation of a set of shapes via descriptor-space spectral clustering. The question inherent to co-analysis is whether we can extract more information by analyzing a set simultaneously, instead of individual or pairs of shapes. Our investigation leads to an affirmative answer as we have shown that the set makes it possible for semantically related parts that differ significantly in geometry and even topology to be linked via third-party connections. By performing the analysis in a descriptor space and exploiting the power of spectral clustering, we are able to properly co-segment sets of shapes exhibiting significant variability.

Ultimately, we would like to subject our co-segmentation approach to large-scale tests. Existing benchmarks for shape segmentation do not quite serve the purpose however. The Princeton Segmentation Benchmark (PSB) [Chen et al. 2009] consists of multiple segmentations per shape that were created individually via automatic methods or manual efforts without taking the sets into account, i.e., they are not necessarily consistent across each shape class. While Kalogerakis et al. [2010] provide a ground-truth labeling of the PSB data which is consistent across the sets, it is not obvious that this is the proper ground-truth to compare to. The ground-truth is prepared to guide the supervised algorithm to what the user desires to extract from the shapes. As Kalogerakis et al. clearly showed in their paper, if an alternative ground-truth is prepared, the algorithm extracts a different segmentation. Hence a comparison to a fixed ground-truth is problematic as well. The preparation of rigorous large-scale tests for co-segmentation of sets of shapes requires significant effort and we leave that for future work.

Nevertheless, we evaluated our approach on a moderately-sized dataset with seven classes of shapes, and compared the co-segmentation to one possible ground-truth prepared by a human. We showed that the result is close to the user-designed segmentations, implying that it does possess semantic meaning.

Limitations. The main limitation of our approach stems from the obvious fact that the quality of the results is entirely dictated by the input set. It may be possible that there is no link between two semantically related segments since the third parties that may establish such a link are missing, e.g., the wired chair and peculiar vases in Figure 7 are unique entities; there are no third parties that relate these shapes to the set. Also, while related parts can be properly linked via the set, imperfect clustering results may cause unrelated segments to be assigned to the same label, e.g., the small candelabra handles and animal parts in Figure 7.

Moreover, the success of our approach is inherently tied to the quality or usefulness of the shape descriptors. Firstly, one limitation of our current descriptors is that their computation requires the input models to be manifold meshes. Secondly, and more importantly, the dependence on shape descriptors leads to the difficult question of what information should be extracted to provide sufficient knowledge about the shapes. Nevertheless, our approach will naturally benefit from incorporating more sophisticated descriptors, e.g., of a structural nature [Biasotti et al. 2008; Shapira et al. 2009]. Another point for practical improvement is the incorporation of more advanced statistical models to represent the clusters of parts.

Future work. It may be naturally expected that our current results do not yet surpass those of supervised approaches that are supported by well-built training sets [Kalogerakis et al. 2010; van Kaick et al. 2011]. However, our set-driven co-segmentation has the potential of outperforming a supervised approach when the knowledge of the latter is insufficient. Perhaps more effective would be a semi-supervised approach, where the user needs to provide only a reduced amount of knowledge and the algorithm can maximally

exploit the power of the input set, or a setting where the user only corrects the parts erroneously labeled so that the system actively learns and adapts the co-segmentation of the set.

The question of how much we can learn from a set is still open for further study, and could lead to novel approaches that take advantage of group information in totally different ways. Another direction for further research is to develop domain-specific shape descriptors, so that the co-analysis can be specialized to specific classes, e.g., humanoid characters, creatures, tools, or vehicles.

Acknowledgements

We thank the anonymous reviewers for their valuable comments and suggestions. Thanks also go to Aleksey Golovinskiy and Thomas Funkhouser for providing their consistent segmentation code, to Evangelos Kalogerakis and collaborators for providing their dataset of labeled shapes and evaluating their supervised method on our dataset, to Xiaobai Chen for the segmentation benchmark, and to Lior Shamir for the SDF code. This research is supported in part by the Israel Science Foundation and the Natural Sciences and Engineering Research Council of Canada (Grant no. 611370 for Hao Zhang and 611393 for Ghassan Hamarneh).

References

- BIASOTTI, S., GIORGI, D., SPAGNUOLO, M., AND FALCIDIENO, B. 2008. Reeb graphs for shape analysis and applications. *Theoretical Computer Science* 392, 1–3, 5–22.
- BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 11, 1222–1239.
- CHEN, X., GOLOVINSKIY, A., , AND FUNKHOUSER, T. 2009. A benchmark for 3D mesh segmentation. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 28, 3, 1–12.
- COIFMAN, R. R., AND LAFON, S. 2006. Diffusion maps. *Applied and Computational Harmonic Analysis* 21, 1, 5–30.
- COMANICIU, D., AND MEER, P. 2002. Mean shift: a robust approach towards feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 5, 603–619.
- DE GOES, F., GOLDENSTEIN, S., AND VELHO, L. 2008. A hierarchical segmentation of articulated bodies. *Computer Graphics Forum (Proc. SGP)* 27, 5, 1349–1356.
- FU, H., COHEN-OR, D., DROR, G., AND SHEFFER, A. 2008. Upright orientation of man-made objects. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 27, 3, 1–8.
- GAL, R., SORKINE, O., MITRA, N. J., AND COHEN-OR, D. 2009. iWIRES: an analyze-and-edit approach to shape manipulation. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 28, 3, 1–10.
- GOLOVINSKIY, A., AND FUNKHOUSER, T. 2009. Consistent segmentation of 3D models. *Computers & Graphics (Proc. of SMI)* 33, 3, 262–269.
- HUANG, Q., KOLTUN, V., AND GUIBAS, L. 2011. Joint shape segmentation with linear programming. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 30, 6.
- JOULIN, A., BACH, F., AND J.PONCE. 2010. Discriminative clustering for image co-segmentation. In *Proc. IEEE Conf. on CVPR*, 1943–1950.
- KALOGERAKIS, E., HERTZMANN, A., AND SINGH, K. 2010. Learning 3D mesh segmentation and labeling. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 29, 3, 1–11.
- KAZHDAN, M., FUNKHOUSER, T., AND RUSINKIEWICZ, S. 2004. Shape matching and anisotropy. *ACM Trans. on Graphics* 23, 3, 623–629.
- MITRA, N. J., YANG, Y.-L., YAN, D.-M., LI, W., AND AGRAWALA, M. 2010. Illustrating how mechanical assemblies work. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 29, 4, 1–12.
- NADLER, B., LAFON, S., COIFMAN, R. R., AND KEVREKIDIS, I. G. 2005. Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In *NIPS*, 1–8.
- ROTHER, C., KOLMOGOROV, V., MINKA, T., AND BLAKE, A. 2006. Cosegmentation of image pairs by histogram matching – incorporating a global constraint into MRFs. In *Proc. IEEE Conf. on CVPR*, 993–1000.
- SHAMIR, A., SHAPIRA, L., AND COHEN-OR, D. 2006. Mesh analysis using geodesic mean-shift. *The Visual Computer* 22, 99–108.
- SHAMIR, A. 2008. A survey on mesh segmentation techniques. *Computer Graphics Forum* 27, 6, 1539–1556.
- SHAPIRA, L., SHALOM, S., SHAMIR, A., COHEN-OR, D., AND ZHANG, H. 2009. Contextual part analogies in 3D objects. *Int. J. Comput. Vision* 89, 2–3, 309–326.
- SIMARI, P., NOWROUZEZHARAI, D., KALOGERAKIS, E., AND SINGH, K. 2009. Multi-objective shape segmentation and labeling. *Computer Graphics Forum (Proc. SGP)* 28, 5, 1415–1425.
- VAN KAICK, O., ZHANG, H., HAMARNEH, G., AND COHEN-OR, D. 2010. A survey on shape correspondence. In *Proc. Eurographics State-of-the-Art Report*, 1–23.
- VAN KAICK, O., TAGLIASACCHI, A., SIDI, O., ZHANG, H., COHEN-OR, D., WOLF, L., AND HAMARNEH, G. 2011. Prior knowledge for part correspondence. *Computer Graphics Forum (Proc. EUROGRAPHICS)* 30, 2, 553–562.
- WANG, Y., XU, K., LI, J., ZHANG, H., SHAMIR, A., LIU, L., CHENG, Z., AND XIONG, Y. 2011. Symmetry hierarchy of man-made objects. *Computer Graphics Forum (Proc. EUROGRAPHICS)* 30, 2, 287–296.
- XU, W., WANG, J., YIN, K., ZHOU, K., VAN DE PANNE, M., CHEN, F., AND GUO, B. 2009. Joint-aware manipulation of deformable models. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 28, 3, 1–9.
- XU, K., LI, H., ZHANG, H., COHEN-OR, D., XIONG, Y., AND CHENG, Z. 2010. Style-content separation by anisotropic part scales. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)* 29, 5, 1–9.
- ZHANG, H., VAN KAICK, O., AND DYER, R. 2010. Spectral mesh processing. *Computer Graphics Forum* 29, 6, 1865–1894.