Unsupervised Co-Segmentation of 3D Shapes via Affinity Aggregation Spectral Clustering

Zizhao Wu[a], Yunhai Wang[†b], Ruyang Shou[a], Baoquan Chen[b], Xinguo Liu[a]

[†] *Corresponding author: cloudseawang@gmail.com*

[a] *State Key Lab of CAD & CG, Zhejiang University*
[b] *Shenzhen VisuCA Key Lab/SIAT*

Abstract

Many shape co-segmentation methods employ multiple descriptors to measure the similarities between parts of a set of shapes in a descriptor space. Different shape descriptors characterize a shape in different aspects. Simply concatenating them into a single vector might greatly degrade the performance of the co-analysis in the presence of irrelevant and redundant information. In this paper, we propose an approach to fuse multiple descriptors for unsupervised co-segmentation of a set of shapes from the same family. Starting from the over-segmentations of shapes, our approach generates the consistent segmentation by performing the spectral clustering in a fused space of shape descriptors. The core of our approach is to seek for an optimal combination of affinity matrices of different descriptors so as to alleviate the impact of unreliable and irrelevant features. More specially, we introduce a local similarity based affinity aggregation spectral clustering algorithm, which assumes the local similarities are more reliable than far-away ones. Experimental results show the efficiency of our approach and improvements over the state-of-the-art algorithms on the benchmark datasets.

Keywords: Co-segmentation, Descriptor fusion, Affinity aggregation, Spectral clustering

1. Introduction

Recently, there has been an increasing interest in shape co-analysis. The premise is that we can extract more knowledge by simultaneously analyzing a set of shapes, rather than an individual shape. The main task of co-analysis is *co-segmentation*, which simultaneously segments all the shapes in the input set in a consistent manner. The consistent segmentation has demonstrated great utility in modeling [1, 2], shape retrieval [3, 4], texturing [5], etc.

Many methods have been designed for shape co-segmentation [5, 6, 7, 8, 9, 10, 11]. In this paper, we focus on unsupervised co-segmentation, where no prior information is given and the entire knowledge must be extracted from the input set. Previous attempts in unsupervised co-segmentation can be classified into alignment-based and descriptor-based. In the alignment-based setting [6, 7], the correspondences among different parts are constructed by a global alignment. Since parts with similar semantics can be rather dissimilar geometrically as well as topologically, these methods cannot handle shape sets with large variations. The descriptor-based methods [8, 9, 11] employ multiple feature descriptors to measure the similarity of parts, and obtain consistent segmentations by clustering the descriptor space. Since the descriptors are independent of the pose and location of shapes, they can handle shapes with rich variations in part composition and geometry.

Different shape descriptors describe different aspects of the geometric characteristics, and often provide complementary information. Figure 1 shows an example, where the geodesic distance to the base of the shape (GB) is more reliable than the shape diameter function (SDF) for discriminating semantic parts in Figure 1 (a,b), while SDF is more reliable than GB in Figure 1 (c,d). Thus, simply concatenating them into a single vector might contain high degree of unreliable, and redundant information. The consistent segmentation obtained by clustering this kind of vectors may be not-optimal. However, with the exception of [9], co-segmentation methods do not address the feature selection or feature weighting in an unsupervised setting.

In this paper, we propose an approach to fuse mul-

tiple descriptors for descriptor-based unsupervised co-segmentation of a set of shapes from the same class. The proposed approach is based on the affinity aggregation spectral clustering (AASC) [12], which extends the spectral clustering to the setting where multiple affinities matrices are available. This algorithm can automatically weight each descriptor properly and aggregate multiple affinity matrices to construct a better one.

By assuming that the local similarities (high values) are more reliable than the far-away ones [13], we construct a sparse affinity matrix based on *k*-nearest neighbor graph. The resulting affinity matrix not only captures the core structure of the feature but also partially removes the unreliability, in comparison to the original one. With that, our local similarity based AASC can effectively combines the strengths of different descriptors, see Figure 2.

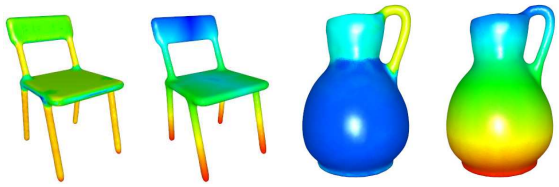
Compared with the subspace clustering approach [9], our approach has two differences. First, our approach explicitly weights each feature and provides an explicit form of the learned affinity, while [9] finds a low dimensional representation of each feature for concentrating all affinity matrices together. Second, our approach integrates the feature selection and spectral clustering into an unified procedure, while [9] works in a two-step procedure. To some extent, our approach is similar with the learning based approach [5], where JointBoost classifier automatically computes the weight of each feature in the classification.

We evaluate our approach on various shape categories and make comparisons with state-of-the-art approaches. The results show that our approach performs better. Since our approach unifies the feature selection and spectral clustering together, it is more efficient than the previous methods. In summary, our contributions are twofold.

- We propose an unsupervised metric weighting method for shape co-segmentation, which simultaneously clusters the descriptor space to generate the consistent segmentation and weights each descriptor.
- We improve the affinity aggregation spectral clustering algorithm with the local similarities, which guarantees the consistency of the co-segmentation results.

2. Related Work

Shape segmentation is a fundamental problem in shape analysis, which refers to decomposing a 3D shape into meaningful parts. Classical shape segmentation approaches focus on finding simple geometric criteria for segmentation of a single input mesh [14]. Although a variety of approaches have been proposed, no segmentation algorithm is known to produce high quality results for all classes of shapes [15]. One reason is that the individual shape may not provide enough geometric cues to distinguish its meaningful parts.



(a) SDF (b) GB (c) SDF (d) GB

Figure 1: SDF and GB defined on a chair and vase, respectively. The back and the seat of the chair are quite similar in SDF (a), while they differ a lot in GB (b); the handle is similar with neck in GB (d), but quite different in SDF (c). Hence, different descriptors for different shape categories should be assigned with different weights.

Recently, researchers have proposed to rather analyze sets of shapes and compute their consistent segmentation. Compared to traditional segmentation approaches, consistent segmentation not only partitions the shapes into segments, but also consistently labels the segments across the set.

Early work by Golovinskiy and Funkhouser [6] builds reliable correspondences across shapes using shape alignment and then clusters the shape faces according to an underlying graph. The graph links faces that are adjacent in the models and faces that establish a correspondence among different meshes after alignment. To deal with non-homogeneous part scales, Xu et al. [7] factor out the scale variation in the shape parts by clustering the shapes into different styles, which is defined as the scales of the shape parts. However, these approaches are limited to the shapes that can be properly aligned.

Kalogerakis et al. [5] present a supervised method to simultaneously segment and label shapes. Given a training set with enough pre-analyzed shapes, a model is learned with hundreds of geometric descriptors and then assigns labels to the new shape based on its descriptors. To enhance the correspondence among parts, van Kaick et al. [16] augment the supervised segmentation with content-driven analysis via a joint labeling

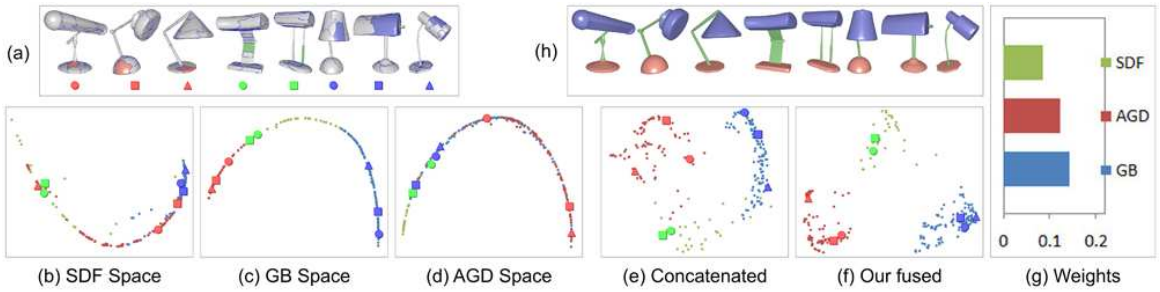


Figure 2: Overview of the steps in our co-segmentation: (a) An over-segmentation is computed for each shape. (b,c,d) The 2D spectral spaces of affinity matrices based on the three computed descriptors, where each patch is corresponding to a point in such spaces. Three parts are mixed together in SDF and AGD spaces (b,d), while they fall in different ranges but not clearly separated in GB space. (f,g,h) Our fused space (f) with different weights for each descriptor (g), where the parts are clearly separated, resulting in the segmentation in (h). (e) In comparison, the 2D spectral space of the affinity matrix of the descriptor generated by concatenating three descriptors together, where the bottom and body of the lamp are mixed.

approach. However, these methods require a substantial number of manually labeled training shapes. This poses a challenge since users are required to manually create training sets

To overcome this limitation, Sidi et al. [8] introduce a descriptor-based method, which poses co-segmentation as a clustering problem in a descriptor space. This method can handle shapes with rich variations in part composition and geometry. Instead of concatenating the different feature descriptors into one vector, Hu et al. [9] formulate the co-segmentation into a subspace clustering problem in multiple features spaces. They first find consistent sparse representation for all descriptors' affinity matrices, and then simply aggregate all matrices into a new one, finally perform clustering to get consistent segmentation. Here, we present a different approach to fuse multiple features. Rather than taking a two-step pipeline, our approach integrates the feature selection and spectral clustering into an unified procedure, and automatically weights each descriptor. Recently, Meng et al. [11] propose to manually weight each descriptor before clustering. However, finding a proper weighting is a difficult problem.

Since the geometry alone cannot always fully convey shape semantics, it is hard to generate a perfect co-segmentation of a set with unsupervised methods. Wang et al. [10] propose a constrained clustering method to co-segment a set of shapes, where the user can actively assist in the clustering process. Since this method is also achieved by clustering the descriptor space, our method can be regarded as complementary to them.

There has been a large body of works in shape analysis using spectral methods [17]. In many applications, there are multiple useful feature descriptors and thereby multiple affinity matrices. However, almost all existing methods concatenate all descriptors into one feature vector and thus perform clustering on a single affinity matrix.

There are a few algorithms which improve the spectral clustering for fusing multiple affinities. Cai et al. [18] propose a multi-modal spectral clustering (MMSC) algorithm, which integrates such heterogeneous features on unlabeled images and unsegmented images. MMSC learns a commonly shared graph Laplacian matrix by unifying different features. Huang et al. [12] propose the affinity aggregated spectral clustering, which aggregates the affinity matrices from different features into an affinity matrix with feature weighting. The authors demonstrate the effectiveness of this algorithm by a variety of clustering problems including image clustering, face clustering and text clustering, in the presence of multiple feature cues. Since the problem of unsupervised co-segmentation can be represented as a clustering problem, we adopt this algorithm to optimally fuse multiple descriptors. In addition, in order to better capture the cluster structure, we construct new affinity matrices based on local similarities to obtain a better clustering result.

3. Overview

Given a set of 3D shapes from the same class, our approach generates a consistent segmentation of a set. Following the similar pipeline with existing methods [9],

our approach consists of three steps. First, we over-segment each shape into many small patches. Then, we compute multiple feature descriptors for each patch. Finally, we perform affinity aggregation spectral clustering over all patches and obtain a co-segmentation of the set. Figure 2 gives an overview of our method, showing at the same time a set of results generated by our algorithm.

Over-segmentation. Similar to some previous works [9, 11], we employ normalized cuts [19] to decompose each shape into primitive patches. The number of patches per shape is set to 30. Then, we use graph cuts [20] to refine the patch boundaries. Figure 2(a) shows several examples of our over-segmentation results.

Feature descriptors. Our co-segmentation is achieved by clustering the descriptor space. Thus, selection of informative descriptors can help us better distinguish patches in different semantic parts.

Here five robust and informative shape descriptors are chosen, i.e., Shape Diameter Function (SDF) [21], Conformal Factor (CF) [22], Shape Contexts (SC) [23, 24], Average Geodesic Distance (AGD) [25], and the geodesic distance to the base of the shape (GB) [8]. All these descriptors are defined on mesh triangles. Taken the computation of SC as an example, given a face, we measured the distribution of all the other faces (weighted by their area) in logarithmic geodesic distance bins and uniform angle bins, where the angles are measured relative to the normal of each face. Therefore, for each face of shapes, there are five scalar values to describe its geometric characteristics. For each descriptor, we compute a histogram to capture its distribution for all faces in each patch. Let p_i denote the i-th patch and $h_{k,i}$ be the histogram for p_i of the k-th descriptor. We define the dissimilarity between two patches p_i and p_j is $d_k(p_i, p_j) = EMD(h_{k,i}, h_{k,j})$, where $EMD(x, y)$ is the earth-mover's distance [26] between x and y. Then, a Gaussian kernel is applied to the distances in order to construct the entries of the affinity matrix for each descriptor:

$$w_{k,ij} = exp(-d_k(p_i, p_j)/2\sigma^2), \tag{1}$$

which represents the similarity distance between the i-th patch and the j-th patch for the k-th descriptor. In our implementation, we set the number of bins to be 50 for each of histograms and σ is set to be the mean of all distances. Figure 3 illustrates the shape descriptors and their corresponding histograms on two examples.

Affinity Aggregation Spectral Clustering. For the five selected descriptors, we construct an affinity matrix for each of them with Equation 1. To better capture the core structure of each descriptor, we refine the affinity matrix based on local similarities. The resulted affinity matrix partially removes the unreliability of the distance measure. Then, we use affinity aggregation spectral clustering to aggregate affinity matrices from different descriptors into a better one. With a formulation of learning the similarity matrix for spectral clustering, our method simultaneously weights each descriptor properly and groups the aggregated affinity matrix to generate the consistent segmentation.

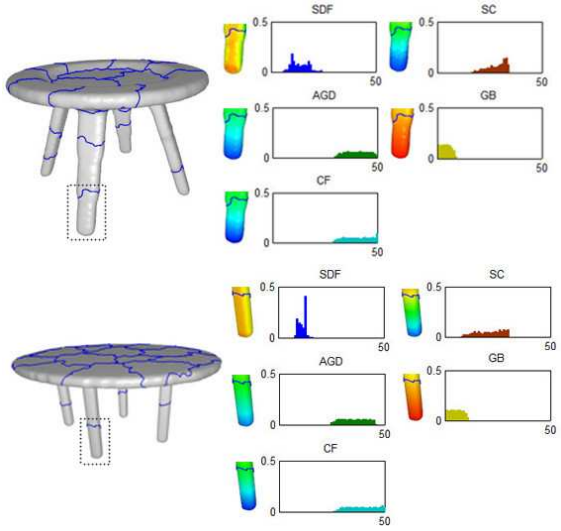


Figure 3: The histograms of five used descriptors of the two tables.

4. Affinity Aggregation Spectral Clustering

We begin by introducing our notation and summarizing the key ideas of spectral clustering. For more details of spectral clustering we refer the reader to von Luxburg [27]. Given n points $x_1, \cdots, x_n$, we define a $n \times n$ affinity matrix W, where each entry W_{ij} encodes the similarity between x_i and x_j.

Spectral clustering partitions the data into m disjoint clusters by finding the indicator vector $F = \{f_1, \cdots, f_n\}$, which satisfies

$$\min_{f_1,\ldots,f_n} \sum_{i,j} W_{k,ij} \| f_i - f_j \|^2 = F'(D - W)F, \tag{2}$$

where D is a diagonal matrix, whose i-th diagonal element is the sum of the elements in the i-th row of W. We adopt the normalized spectral clustering [28] to perform clustering. Taking the constraint $F'DF = 1$, the indicator vector F can be obtained by finding the eigenvectors

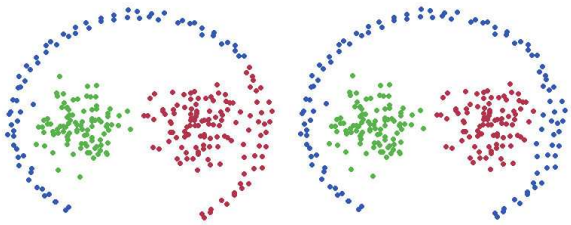


Figure 4: (Left) Spectral clustering based on the full connected graph. (Right) Spectral clustering based on the k-nearest neighbor graph where k is 5.

u and eigenvalues λ of the generalized eigenvalue system $Lu = \lambda Du$, where L is the Laplacian matrix of W.

4.1. Local Similarities

In Equation 2, W is constructed by computing the pairwise similarities between any two points on a fully connected graph G. The patches generated by the initial over-segmentation in Section 3 have a large amount of variability. Thus, the cluster structures in W may be destroyed by noise, which are introduced by some small patches.

By assuming that the local similarities are more reliable than far-away ones [13], we construct another graph $\tilde{G}$ to remove the unreliable neighbors. The vertices of G are the same as in $\tilde{G}$, and the edges are the nearby ones in G. In other words, we select k nearest neighbors for each vertex. Using k-nearest neighbors, we construct $\tilde{G}$ with associated similarities matrix $\tilde{W}$:

$$\tilde{W}_{k,ij} = \begin{cases} W_{k,ij}, & p_j \in knn(p_i) \text{ or } p_i \in knn(p_j) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where knn denotes k-nearest neighbors of the point p_i. Figure 4 shows an example, where our local similarities based spectral clustering performs better than the one with full connected graph.

4.2. Affinity Aggregation

Now there are m descriptors, and each of them has an affinity matrix W_k. As shown in Figure 1, different descriptors should be assigned different weights. Here, we use the affinity aggregation spectral clustering (AASC) [12] to automatically find an appropriate weight for each descriptor. Let $\alpha = \{\alpha_1, \ldots, \alpha_m\}$ be the weights associated with the affinity matrices. AASC can be formulated as:

$$\begin{aligned} & \min_{\alpha,F} \textstyle\sum_{k=1}^{m} \sum_{i,j} \alpha_k w_{k,ij} \|f_i - f_j\|^2 \\ = & \min_{\alpha,F} \textstyle\sum_{k=1}^{m} \alpha_k F'(D_k - W_k)F \quad (4) \\ = & \min_{\alpha,F} \textstyle\sum_{k=1}^{m} \alpha_k r_k, \quad (5) \end{aligned}$$

where D_k is a diagonal matrix of W_k which is constructed the same way as $\tilde{D}_k$, $D_k - W_k$ is the Laplacian matrix of the k-th descriptor and

$$r_k = F'(D_k - W_k)F.$$

Compared to Equation 2, the weight of each descriptor is also incorporated into this new objective function. With a weighted combination of the affinity matrices, the clusters can be better separated as shown in Figure 2.

This minimization involves two sets of unknown variables: α and F and is generally hard to solve. We design a two-step approach to approximate the minimization by alternatively fixing α and F. When the weights α are fixed, F can be found by the standard spectral clustering. When F is fixed, we solve α by exploiting two additional constraints.

The first constraint comes from the normalized spectral clustering, i.e., the final diagonal matrix D must satisfy the following equation:

$$1 = F'DF = F'(\alpha_1 D_1 + \cdots + \alpha_m D_m)F = \sum_{k=1}^{m} \alpha_k s_k, \quad (6)$$

where

$$s_k = F' D_k F.$$

With this constraint, the spectral clustering usually converges to a nice clustering result [29].

Since the diagonals of the affinity matrices are all 1s, and all elements of these matrices are non-negative, we denote $w_{k,ij} = v_{k,ij}^2$ and V_k denote the matrix constituted of $v_{k,ij}$. The second constraint comes from the sum of the weighted matrices in a normalized condition, i.e.,

$$tr(\sum_{k=1}^{m} \sqrt{\alpha_k} V_k) = n, \quad (7)$$

where tr is the *trace* of a matrix. This constraint implies that the traces of such matrices are bounded.

We introduce the above constraint because we prefer a lower bound on the trace of the weighted matrices. Specifically, if the trace of the weighted matrix is lowly bounded, the sum of its eigenvalues is also lowly

bounded. This could prevent the eigenvalues approaching very small positive values so as to guarantee a stable clustering result.

By Cauchy-Schwartz inequality, we have $\sum_{k=1}^{m} tr(\alpha_k W_k) \geq \frac{1}{m}(\sum_{k=1}^{m} \sqrt{\alpha_k} tr(V_k))^2 = \frac{n^2}{m}$. Since $tr(V_k) = n$, the second constraint becomes:

$$\sum_{k=1}^{m} \sqrt{\alpha_k} = 1. \quad (8)$$

By applying Lagrange multiplier to these two constraints, the problem of finding the proper α can be reduced to a 1-D line search problem, which is easy to solve. For more details, we refer the reader to [12].

In general, AASC can be solved with a two-step iterative algorithm. Specifically, starting with $\alpha_k = 1/m$, we alternate between finding F with normalized spectral clustering and searching optimal weights α with 1-D searching under both constraints (Equation 6 and Equation 8), keep reducing the error and ensuring the convergence of the iterative process. Hence, the time complexity of AASC is $O(n^3 * m)$, where n is the number of points and m is the number of iterations. After obtaining F, we run k-means in the indicator space to cluster the data into C classes.

4.3. Affinity Aggregation based Co-segmentation

Given N shapes from the same class, our goal is to consistently segment each shape into K meaningful parts. As mentioned in Section 3, we first over-segment each shape into P patches and then extract H descriptors from each patch and compute a $NP \times NP$ affinity matrix $\tilde{W}_i$ for each descriptor.

After obtaining affinity matrices $\tilde{W}_1, \ldots, \tilde{W}_H$, there are two ways to apply AASC. First, we can directly apply AASC on them and get the results as shown in the middle of Figure 5, which are not as good as we expect. We conjecture that the reason is that there is much noise in $\tilde{W}_i$. Thus, before applying AASC, we first construct local similarity based affinity matrices $W_1, \ldots, W_H$ according to Equation 3. The right of Figure 5 shows the results, where the labeling results are better than the ones in the left and middle of Figure 5. The accuracy of the table set is close to error free, and the accuracy of the human is 78.0%, which are better than previous methods [9] (70.4%). Unless otherwise specified, all results in this paper are generated by this method of local similarity based AASC.

For comparison, we simply concatenate all descriptors into one and then apply spectral clustering to the affinity matrix of this new descriptor. On the left of Figure 5, we can see the result generated by this manner, where almost all shapes are mislabeled especially the human set. These results are close to what we expect, because it assigns equal weight to each descriptor.

Table 1: Statistical evaluation of the average accuracy by our algorithm on all categories.

Category	Accuracy	Category	Accuracy
Candelabra	97.2	Chair	87.6
Fourleg	87.7	Goblet	98.7
Guitar	97.4	Lamp	98.4
Vase	83.4	Table	98.4
Human	78.0	Cup	90.2
Iron	79.4	Tele-alien	75.3
Large vase	87.5	Large Chair	90.6

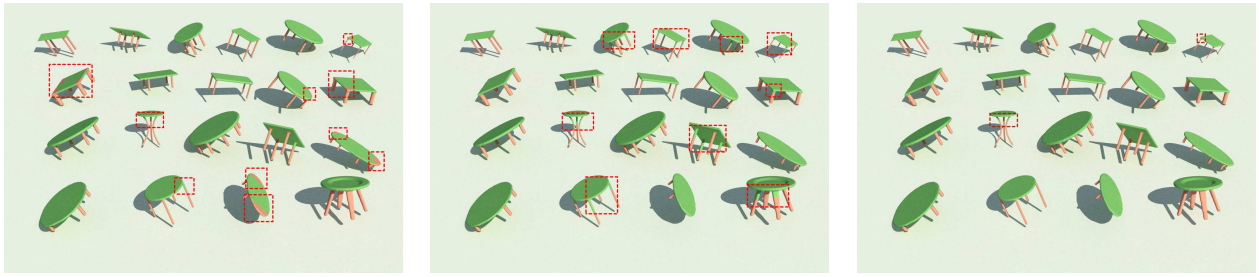
5. Results

In this section, we present our experimental results with benchmark datasets, and compare our method with some state-of-the-art methods.

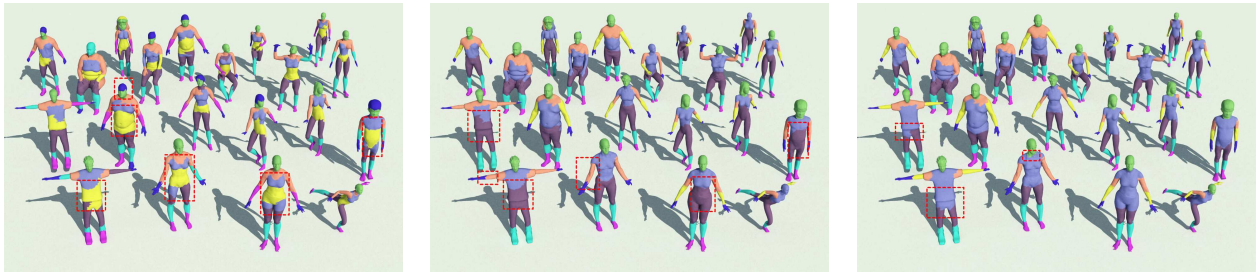
Experimental Dataset. We test our approach on a shape co-segmentation dataset (COSEG) [10], which includes a consistent ground truth segmentation of 8 small sets and 3 large sets. In addition, we select 3 another categories (Human, Table and Cup) with upright orientation from the Princeton Shape Benchmark (PSB) [15], and use the ground truth based on the labeled database [5]. For the ground truth with different number of semantic parts, we manually merged some labels according to the number of the desired clusters of the categories.

Co-segmentation results. There are only two parameters in our approach: the number of neighbors (k) in computing the local similarity and the number of clusters. All results were generated with $k = 20$. Figure 5, 6, 7 show the co-segmentation results of our approach on various categories. From these results, we can see that our approach is able to consistently segment most of shapes, despite the variety in the shape parts from each of the categories.

Because the used descriptors can well characterize the man-made objects, our approach does work well for the Candelabra, Goblet, etc, see Figure 6. Although the shapes in the Vase set have various topologies and low geometric similarities among compounding parts, our approach is able to identify the bases, the bodies and the handles. For the articulated categories like Human and Fourleg, our approach also works well, although there are large variations in pose.



(a) Table



(b) Human

Figure 5: We compare our approach with another two fusion strategies on the sets of Table and Human. (Left) The result generated by first concatenating all descriptors into one descriptor vector and then applying spectral clustering on the local similarity based affinity matrix. (Middle) The result generated by applying AASC on the affinity matrices, which are constructed on the full connected graph. (Right) The result generated by applying AASC on the local similarities based affinity matrices, which are constructed on the k-nearest neighbor graph. The parts within the red box are mislabeled.

Figure 6 and Figure 7 visualize the weights of the descriptors for each set. From the bar charts, we can see that the weights of different descriptors are different but the differences are not large. Almost all weights fall in the range of 0.03 to 0.05. The reason is that these descriptors are correlated with each other. For example, GB and AGD both characterize the shape with the geodesic distance.

Except the sets of Iron and Cup, SC and AGD are ranked as top 3 descriptors for the left categories. This is also verified by [5]. For the parts with different thickness, such as the sets of Vase and Cup, SDF does work well, but performs poorly for the Chair, since the thickness of the middle part in the back is close to that of the leg. This is also true for the Human set, where the thickness of the arm is close to that of the leg. For the shape with upright orientation, GB plays an important role, as shown by results of the Vase and Chair sets in Figure 7.

Evaluation. To evaluate the quality of our results, we use the accuracy measure defined in [10] using the correctly labeled shape area:

$$Accuracy(l,t) = \sum_i a_i \delta(l_i = t_i) / \sum_i a_i, \qquad (9)$$

where a_i is the area of face i, l is the label computed by the co-segmentation, t is the ground-truth labeling, α_i is the area of face i, and $\delta(x = y)$ takes value 1 if and only if x equals y. We then average the accuracies for all the shapes in the set to obtain the accuracy of the

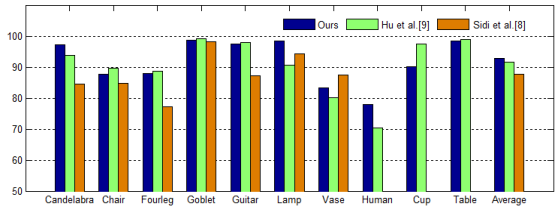
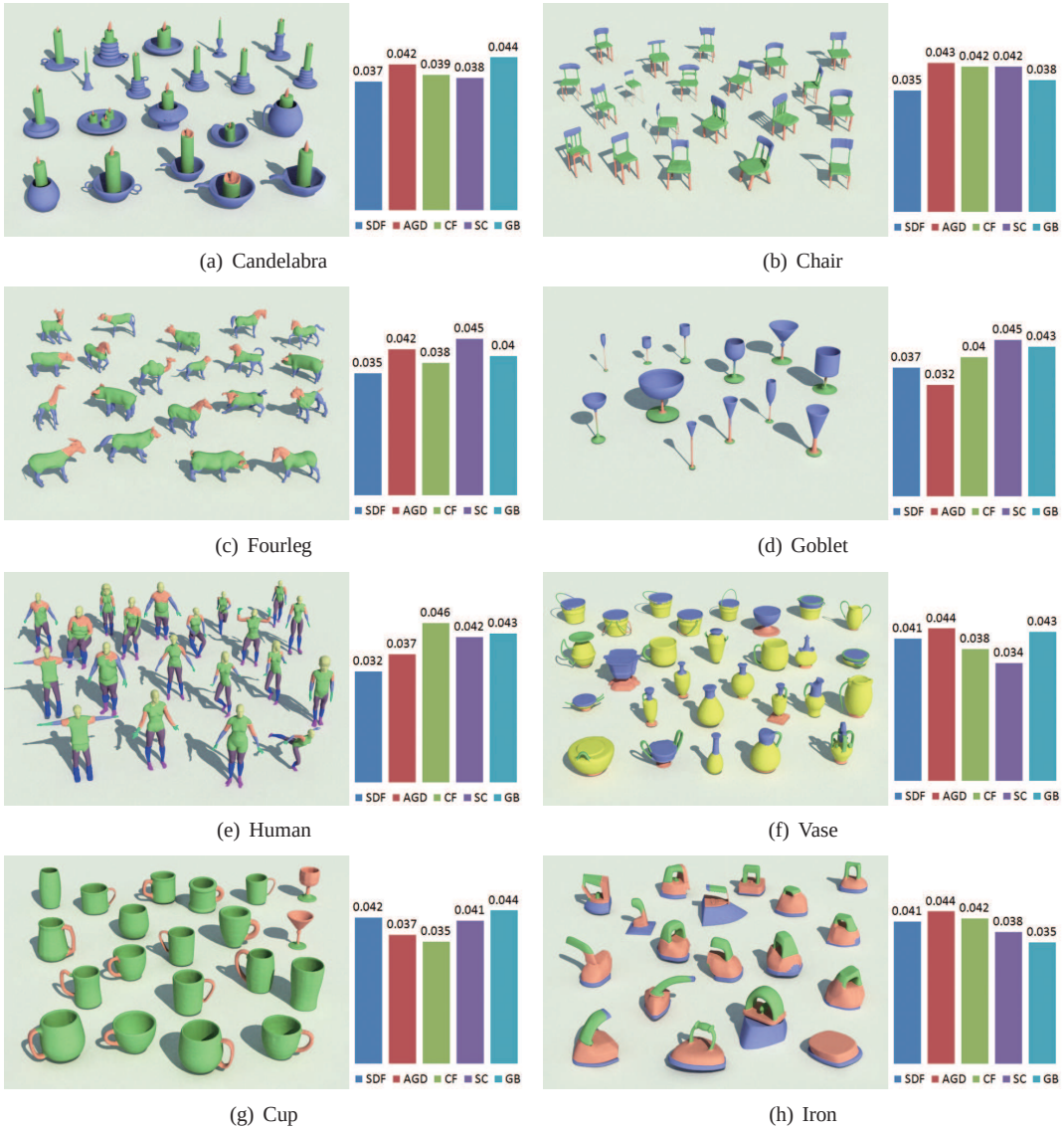


Figure 8: Comparison between our approach with two state-of-the-art techniques [8] and [9] on some categories. Higher values indicate higher accuracies.



(a) Candelabra (b) Chair

(c) Fourleg (d) Goblet

(e) Human (f) Vase

(g) Cup (h) Iron

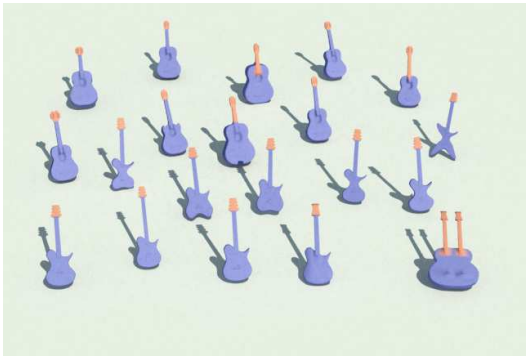
Figure 6: Results of our co-segmentation on the COSEG dataset [10] and the PSB dataset [15]. Corresponding segments are shown with the same color. Statistical evaluation for categories are shown in Table 1.

co-segmentation method. Table 1 shows the statistical evaluation results, where our algorithm has obtained an average accuracy of 89.3% over all categories.

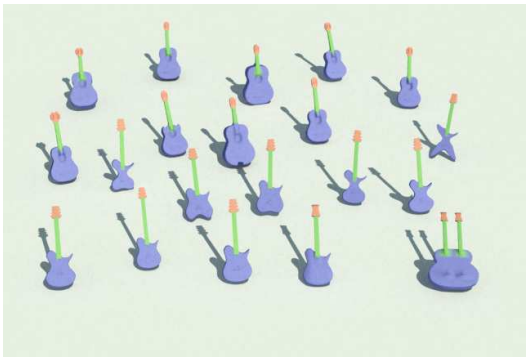
The three categories having lowest average accuracies are Tele-alien, Human and Iron. The first two categories are non-rigid deformable shapes, whereas the descriptors used in this paper are aimed for upright orientable shapes. As for the set of Iron, the shape variation is too large to well distinguish the parts.

When the size of the set becomes larger, our method tends to generate better results. From Table 1, we can see that the accuracy on the large sets of Vase and Chair are improved 3%-4% compared to the corresponding small sets.

Comparisons to state-of-the-art. We make a comparison to two state-of-the-art works: the unsupervised approach using concatenating descriptors [8] and the unsupervised method using subspace clustering [9]. Figure 8 shows the comparison of results regarding the co-segmentation accuracies among these three methods.

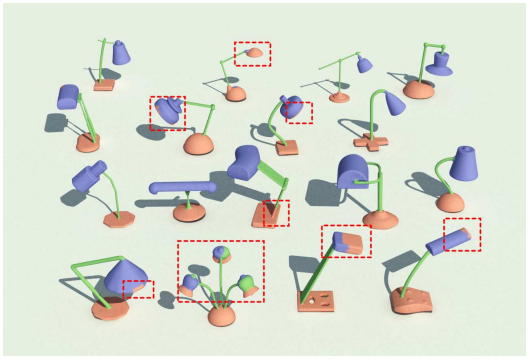


(a) Sidi et al. [8]

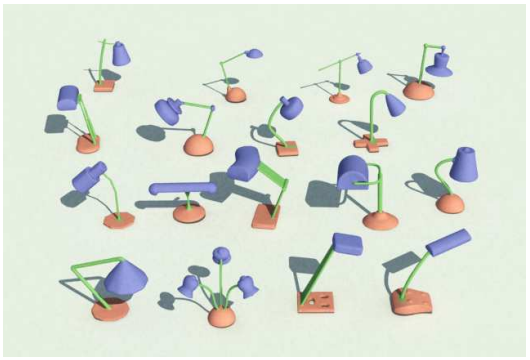


(b) Ours

Figure 9: Co-segmentation results of Sidi et al. [8] and ours on the Guitar dataset.



(a) Hu et al. [9]



(b) Ours

Figure 10: Co-segmentation results of Hu et al. [9] and ours on the Lamp dataset.

In Figure 8, we can see that the average accuracies of [8] and ours are 87.7% and 92.9%, respectively. Our results get higher accuracy than theirs for most of categories. Figure 9 shows some results by applying these two approaches on the Guitar set. We can see that their approach failed to separate the head and neck parts, whereas ours can distinguish them well. This comparisons shows that our descriptor fusion is more efficient than simple concatenating features in dealing with multiple descriptors.

Compared to the work of Hu et al.[9], our result is slightly better than theirs, where the average accuracies are 91.7% vs. 90.7%. For the sets of Lamp and Human, our result is better than theirs. Figure 10 shows the result of Lamp set, where the method in [9] cannot accurately identify the boundary of the top and takes some tops as the bottom. One reason is that there is noise in the initial affinity matrices due to large shape variations. Compared to [9], our local similarity based AASC can alleviate this problem. However, our approach performs worse on the Cup set (see Figure 6(g)), arising due to the poor initial over-segmentation.

Performance. We evaluated the performance of our approach on a 2.83GHz Intel Core™ 2 Quad processor, 4GB of RAM. The execution time of our pipeline mainly depends on the computation of descriptors, particularly the computation of geodesic distance and shape context, while the other steps are fast. Since the number of superface for each shape is 30, AASC can be finished in less than 1 min within 20 iterations given 20 shapes. One iteration of metric weighting takes on average less than 1 second for eigen-decomposition and 1-D optimization. In all, it takes 8 minutes to go through the whole pipeline for small datasets (20 shapes), and around 30 minutes for large datasets (400 shapes).

6. Conclusions

In this paper, we presented a method to fuse multiple descriptors for unsupervised shape co-segmentation. Our main contribution is to simultaneously cluster the descriptor space and weight each descriptor with the

aid of affinity aggregation spectral clustering. We also showed that local similarity based affinity aggregation is more robust to guarantee the consistency of co-segmentation results. We demonstrated the effectiveness of our approach on several benchmark datasets. The comparison between our results and some state-of-the-art techniques shows that our approach is more effective.

Limitations and future work. There are many limitations of our algorithm, which suggest many avenues for future work. First, our approach suffers from lacking of an automatic way to determine the clustering number which is now specified manually. In the future, we plan to incorporate our feature weighting into self-tuning spectral clustering [30] to resolve this issue. Second, although our fusion method can effectively weight different descriptors, it does not work if none of a good descriptor can properly characterize the shape, since our approach is purely driven by the shape descriptors. For example, many robust descriptors, e.g. AGD and SDF, do not work for non-manifold shapes. Thus, we plan to explore more inherent information to analyze a set of shapes. Third, our current affinity aggregation method can still be improved. For example, fusing multiple descriptors through diffusion process can further improve the aggregation performance [31]. Finally, We hope that the idea of fusing multiple descriptors has a great potential for other applications.

Acknowledgment

Many thanks to the anonymous reviewers for their valuable comments. This work was supported by NSFC (NO.61202222 and No.60970074), China 973 Program (No. 2009CB320801), Fok Ying Tung Education Foundation, and the Fundamental Research Funds for the Central Universities.

References

[1] T. A. Funkhouser, M. M. Kazhdan, P. Shilane, P. Min, W. Kiefer, A. Tal, S. Rusinkiewicz, D. P. Dobkin, Modeling by example, ACM Transactions on Graphics (Proc. SIGGRAPH) 23 (3) (2004) 652–663.

[2] K. Xu, H. Zhang, D. Cohen-Or, B. Chen, Fit and diverse: Set evolution for inspiring 3d shape galleries, ACM Transactions on Graphics (Proc. SIGGRAPH) 31 (4) (2012) 57:1–57:10.

[3] S. Shalom, L. Shapira, A. Shamir, D. Cohen-Or, Part analogies in sets of objects, in: 3DOR, 2008, pp. 33–40.

[4] S. Chaudhuri, E. Kalogerakis, L. J. Guibas, V. Koltun, Probabilistic reasoning for assembly-based 3d modeling, ACM Transactions on Graphics (Proc. SIGGRAPH) 30 (4).

[5] E. Kalogerakis, A. Hertzmann, K. Singh, Learning 3d mesh segmentation and labeling, ACM Transactions on Graphics (Proc. SIGGRAPH) 29 (4) (2010) 1–11.

[6] A. Golovinskiy, T. A. Funkhouser, Consistent segmentation of 3d models, Computers & Graphics 33 (3) (2009) 262–269.

[7] K. Xu, H. Li, H. Zhang, D. Cohen-Or, Y. Xiong, Z. Cheng, Style-content separation by anisotropic part scales, ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 29 (5) (2010) 184:1–184:10.

[8] O. Sidi, O. van Kaick, Y. Kleiman, H. Zhang, D. Cohen-Or, Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering, ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 30 (6) (2011) 126:1–126:9.

[9] R. Hu, L. Fan, L. Liu, Co-segmentation of 3d shapes via subspace clustering, Computer Graphics Forum (Proc. SGP) 31 (5) (2012) 1703–1713.

[10] Y. Wang, S. Asafi, O. van Kaick, H. Zhang, D. Cohen-Or, B. Chen, Active co-analysis of a set of shapes, ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 31 (6) (2012) 165–174.

[11] M. Meng, J. Xia, J. Luo, Y. He, Unsupervised co-segmentation for 3d shapes using iterative multi-label optimization, Computer-Aided Design 45 (2) (2013) 312–320.

[12] H.-C. Huang, Y.-Y. Chuang, C.-S. Chen, Affinity aggregation for spectral clustering, in: Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 773–780.

[13] R. R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5–30.

[14] A. Shamir, A survey on mesh segmentation techniques, Computer Graphics Forum 27 (6) (2008) 1539–1556.

[15] X. Chen, A. Golovinskiy, T. Funkhouser, A benchmark for 3D mesh segmentation, ACM Transactions on Graphics (Proc. SIGGRAPH) 28 (3) (2009) 1–12.

[16] O. van Kaick, A. Tagliasacchi, O. Sidi, H. Zhang, D. Cohen-Or, L. Wolf, G. Hamarneh, Prior knowledge for shape correspondence, Computer Graphics Forum (Special Issue of Eurographics) 30 (2) (2011) 553–562.

[17] H. Zhang, O. Van Kaick, R. Dyer, Spectral mesh processing, Computer graphics forum 29 (6) (2010) 1865–1894.

[18] X. Cai, F. Nie, H. Huang, F. Kamangar, Heterogeneous image feature integration via multi-modal spectral clustering, in: Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1977–1984.

[19] A. Golovinskiy, T. A. Funkhouser, Randomized cuts for 3d mesh analysis, ACM Transactions on Graphics (Proc. SIGGRAPH) 27 (5) (2008) 1–12.

[20] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, in: ICCV, 1999, pp. 377–384.

[21] L. Shapira, S. Shalom, A. Shamir, D. Cohen-Or, H. Zhang, Contextual part analogies in 3d objects, International Journal on Computer Vision 89 (1-2) (2010) 309–326.

[22] M. Ben-Chen, C. Gotsman, Characterizing shape using conformal factors, in: 3DOR, 2008, pp. 1–8.

[23] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.

[24] M. Körtgen, G.-J. Park, M. Novotni, R. Klein, 3d shape matching with 3d shape contexts, in: The 7th central European seminar on computer graphics, Vol. 3, 2003, pp. 5–17.

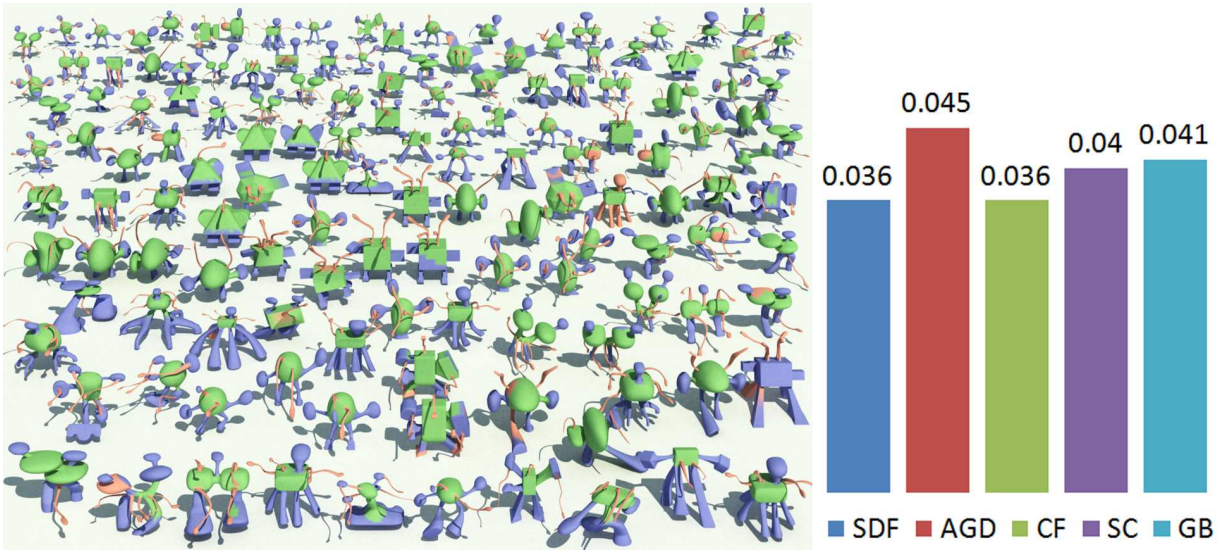
[25] M. Hilaga, Y. Shinagawa, T. Komura, T. L. Kunii, Topology matching for fully automatic similarity estimation of 3d shapes, in: SIGGRAPH, 2001, pp. 203–212.

[26] Y. Rubner, C. Tomasi, L. J. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision 40 (2) (2000) 99–121.

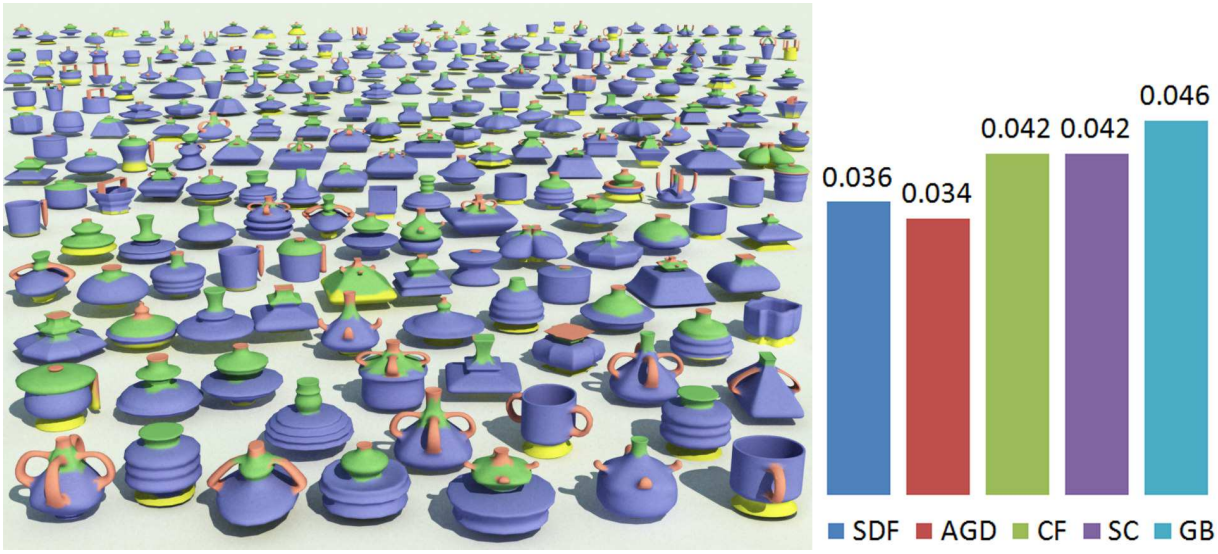
[27] U. Von Luxburg, A tutorial on spectral clustering, Statistics and computing 17 (4) (2007) 395–416.

[28] J. Shi, J. Malik, Normalized cuts and image segmentation,

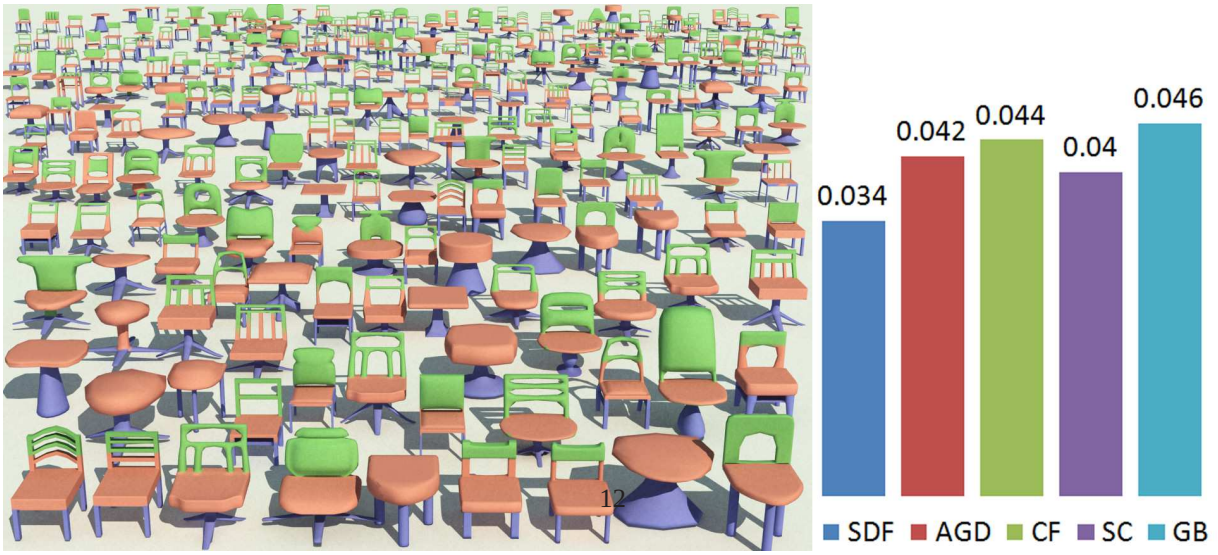
IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905.
[29] U. von Luxburg, A tutorial on spectral clustering, Statistics and Computing 17 (4) (2007) 395–416.
[30] P. Perona, L. Zelnik-Manor, Self-tuning spectral clustering, in: Advances in neural information processing systems'04, 2004, pp. 1601–1608.
[31] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, Z. Tu, Unsupervised metric fusion by cross diffusion, in: Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 2997–3004.



(a) Large Tele-alien



(b) Large Vase



(c) Large Chair

Figure 7: Results of our co-segmentation on the large datasets of COSEG [10].