Hierarchical Surface Abstraction Using Adaptive Mean Shift

Chang Liu, Jiaxi Hu, Jing Hua *
Wayne State University

Hong Qin
Stony Brook University

Abstract

This paper presents a novel hierarchical surface abstraction technique based on Mean Shift for 3D shape matching. The automatically generated statistical modes and their associated geometric properties from Mean Shift-based surface analysis are further integrated into an Attributed Relational Graph (ARG). Therefore, the ARG can be used as an abstract representation of the 3D surface object for 3D shape matching, which has direct applications on motion synthesis, capturing and transfer. In particular, we propose an adaptive mean shift technique and a hierarchical framework based on normalized cuts. Our experiments show that the surface abstraction technique is robust, insensitive to resolution changes and stable to noise. We have tested the technique in 3D shape matching. The experimental results demonstrate the effectiveness of the presented method.

Keywords: Adaptive Mean Shift, Attributed Relational Graph, Shape Matching

1 Introduction

The rapid development of 3D object digitalization equipments and ever increasing demand of shape animation applications have driven the pertinent shape analysis and matching to become a very active research field in computer graphics. Shape matching becomes increasingly popular in many animation applications in the past few years. For example, it was used in motion synthesis [1] and motion capture [2]. Tangelder et al. classified all existing 3D shape retrieval methods into several categories [3]. It was shown that even though global feature-based methods are more efficient, only algorithms based on local features can support accurate detailed matching. The prior work again raises the question of how to effectively integrate local and global features in a single framework. The solution to this question will certainly lead to a better understanding of a given shape. To facilitate most shape comparison tasks, global appearance, local variance and topological features of shapes should be considered together in a systematic way. Taking efficiency into consideration, redundancies should be hidden from execution. Therefore, an automatic abstraction algorithm is a very appealing solution for shape description, which can retain both global and local shape characteristics.

In most cases, a complex shape is composed of relatively simple components. Recent human perception study proved that visual content is recognized by components or parts [4, 5]. We observed that, on one hand, the global feature of a shape can be treated as the composition of the features from its components, on the other hand, partial matching is usually documented as the task of identifying non-trivial, similar components of more complex shapes. These components either have similar semantical meaning from a global point of view or simply share similar local properties. Overall, it requires a systematic approach to represent global composition and local feature information of the components properly. In this paper, we present a hierarchical surface abstraction framework based on adaptive Mean Shift for au-

*Correspondence to: J. Hua. Email: jinghua@wayne.edu

tomatic construction of such a shape representation. Through hierarchical abstraction, we can describe the shape in different levels-of-details. At each level, we only focus on a specific scale of components. This hierarchical abstraction approach is an ideal way to organize shape characteristics towards a representation for all-purpose retrieval tasks. Particularly in this paper, we focus on the surface objects. The surface abstraction is achieved through an adaptive mean shift-based decomposition and a graph-based organization, which together form a powerful description of the shape.

The main contributions of this paper are summarized as follow:

1. We propose a novel adaptive manifold mean shift approach to automatically decompose the surface into meaningful subregions, which are called Feature Enriched Components (FECs), based on its global and local statistics in the feature space. Since geometry and topology information can both be encoded into the feature space, FEC decomposition is a more general way to characterize the shape. It can handle surfaces of complex geometry and arbitrary topology.

2. We present a hierarchical surface abstraction framework. The surface subregions are hierarchically organized into an attributed relational graph (ARG) which reflects both global structures and local levels-of-details of the shape in terms of topology and geometry. The compact shape description facilitates efficient matching and comparison.

3. The shape representation enables many different shape matching and retrieval tasks. Both the partial matching and deformable model retrieval can be achieved using FECs and ARG. Objects with high-genus shapes can be handled without extra efforts.

The rest of the paper is organized as follows. Section 2 reviews the major work that is related to our approaches. In Section 3, we present the adaptive mean shift method and demonstrate how to use it to decompose a surface shape represented by triangular mesh into FECs. In Section 4, we explain how to use an attributed relational graph based representation to achieve global and local abstraction of a shape. In Section 5, we show the results of different shape retrieval applications based on our techniques. Finally, in Section 6, we discuss the advantages and disadvantages of our framework, and potential future work.

2 Related work

Our work is motivated by the state of art in the shape matching field. This section reviews some most related work.

The basic operation of partial matching is to match similar geometries between a query input object and samples in the database. Therefore, the subregions of the sample must be properly defined. To avoid exhaustive search, some research work focused partial matching on "salient" region detection and comparison. One of the typical work was done by Gal and Cohen-Or [6]. They tried to find subregions that are "salient" based on local curvature variances. However, without considerable modification, it may not give the same discriminative power when features other than curvatures are considered. As an alternative, Shilane et al. [7] focused on finding distinctive regions on 3D surfaces. Unlike the salient region method, which measures how much a region sticks out from the rest of the object rather than how important the region is for defining the object type, the distinctive regions are obtained by searching for unique regions among the whole query set. The main problem of this method is its efficiency, the distinctive regions need to be recomputed when the query set is changed. Some other work proposed to extract salient points rather than subregions [8, 9]. Although these points can also take properties such as scales [10], they seem to be too local and not semantic enough for matching purposes. And they are error prone without careful local feature calculation.

Deformable shape retrieval methods are usually skeleton based or graph-based [11, 12, 13]. Although these approaches achieved certain success when models with similar skeletons are considered as in the same class, the skeleton is good at representing shape topology but not at describing surface geometry, thus, they are not accurate for detail matching. Furthermore, the nodes defined on the skeletons are too local to have enough representative power, thus intraclass dissimilarity can not be measured precisely. The idea of combining both topological and geometric surface representations of a shape had motivated Gary et al. to build a "surface skeleton" based on Level Set Diagram (LSD) [14]. In their implementation, the nodes are defined as surface subregions, which form an atlas of the surface. Geometry features are extracted from these subregions, so that they are more suitable for intraclass comparison. However, the algorithm can only handle genus-zero manifold, and may perform poorly when noise is present.

To overcome the difficulties of previous methods, we developed a method to extract FECs of a shape, which can support both partial and deformable model retrieval.

3 Surface decomposition based on adaptive mean shift

We believe that the existing component-based shape representations are not suitable for universal tasks, since the component definitions are either not simple enough to be stable, or not having enough discriminative power.

As a summery, the components should have at least two properties to form a good shape representation for common applications such as matching:

1. The component should not be too small or too local. Otherwise, it is statistically unstable, and not semantically meaningful. The component should not contain a large and complex region either.

2. The samples inside a component should share the same geometrical or physical features, thus, the components can be easily found by clustering. This also implies certain simplicities of the components.

These two properties ensure strong discriminative power, good robustness, and high efficiency.

Although in our implementation, the decomposition is base on general geometric features, such as curvature and normal, other features can be easily embedded to make the patches feature enriched. Thus, we name components satisfying the above two properties as feature enriched components(FECs). A modified mean shift algorithm is used to extract FECs from a shape.

3.1 Manifold mean shift and mode seeking on manifold

Mean shift is a robust approach for feature space analysis [15, 16]. It is widely applied in computer vision applications such as filtering ,segmentation and tracking. It is a statistical approach which is independent of resolution and noise competing. Shamir et al. in [17] extended the mean shift method for feature space analysis of triangular meshes.

The basic idea of mean shift is to find the statistical modes of the sample data, which are usually determined in a high dimensional feature space. The probability density function (PDF) of the data is estimated using a multivariate kernel density estimator:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K(x - x_i), \qquad (1)$$

where K is the d-variate kernel which usually takes the form of $K(x) = c_{k,d}k(\|x\|^2)$, where $c_{k,d}$ is a normalization constant.

To find the mode of each sample, in each step, the current sample point moves toward the mean value of a certain neighborhood along the gradient direction. The offset is calculated as:

$$\delta(x) = \frac{\sum_{i=1}^{n} x_i g(\|\frac{x-x_i}{h}\|^2)}{\sum_{i=1}^{n} g(\|\frac{x-x_i}{h}\|^2)} - x, \qquad (2)$$

where $g(x) = -k'(x)$ and h is the bandwidth. In [15], Comaniciu and Meer further modified the d-variate kernel to be $K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p} k(\|\frac{x^s}{h_s}\|^2) k(\|\frac{x^r}{h_r}\|^2)$, so that the spatial part x^s (stands for locations) can be separated from the range part x^r (stands for properties) for respective controls.

Due to the sound robustness of mean shift, Shamir et.al had extended this framework onto triangular meshes [17]. However, directly move the metric for meanshift calculation onto a closed manifold will cause bad localization of the modes, and generate sibling modes 1.

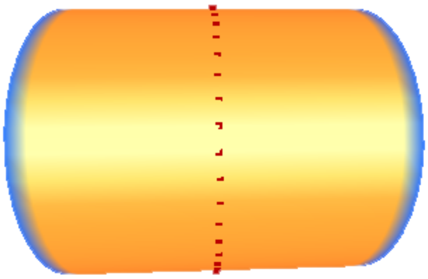

Figure 1: Sibling modes detected by manifold mean shift: the cylinder is colored by mean curvature, multiple modes are found on its mother lines, these modes represent the same cluster.

Furthermore, kernel size should be determined wisely since neighborhood query becomes expensive on a nonparametric surface.

To solve the above problems, we present a two-stage approach based on our novel adaptive mean shift, and the normalized cut.

3.2 *Adaptive manifold mean shift for shape decomposition*

The adaptive manifold mean shift is designed to find the best kernel size for each sample(vertex) as follows:

1. Run manifold mean shift with a small window size w_{MIN}, which determines the smallest possible clusters in the final atlas. Set the repeatability r of each mode to be 1.

2. Increase the window size by δw and center windows at the modes found in step 1. Calculate the stability (will be discussed later) of the mode under the current kernel setting. If the corresponding mode is still a "basin of attraction", ie. an extremum, increase its repeatability r by 1.

3. Repeat step 2 until the mode is not stable any more, or the window size reaches the predefined maximum w_{MAX}, which determines the largest possible clusters. For each mode, store the maximum window size w_{max} under which the mode is still stable.

In order to decide the stability of a mode given a certain truncated kernel function, two terms are calculated:

1. The mean shift offset $\delta(m)$ from equation 2 by centering the window at the mode.

2. The estimated probability density $f(m)$ calculated by equation 1.

If the mode is still stable, $\delta(m)$ should not be too large and $f(m)$ shouldn't be too small. The decision is made by thresholding both of them.

After the above stability test, each vertex will be covered by windows with sizes, $dw^1_{max}, dw^2_{max}, \cdots$, and dw^p_{max}, determined by adjacent modes $m_1, m_2, \cdots, m_p$. The largest window size is selected at the vertex for its mean shift calculation. In this first stage, the initial FECs are obtained by grouping vertices converging to the same mode.

3.3 *Second stage clustering and hierarchical FEC decomposition*

To deal with the bad localization of modes, a second stage clustering is carried out. After the adaptive manifold mean shift, we further cluster the modes according to their Euclidian distances between each other. This problem could be easily transformed to a graph partition problem, which has a robust solution by normalized cut [18]. Given an initial FEC decomposition with M clusters represented by M' modes, a weight matrix W for normalized cut is constructed as follows:

$$\begin{aligned} &W = \{w_{i,j} \mid i = 0, 1, \cdots, M-1; j = 0, 1, \\ &\cdots, M-1\} \\ &w_{i,j} = exp(-dis(C_i, C_j)), \end{aligned} \tag{3}$$

where $dis(m_i, m_j)$ is the average euclidian distance between modes in cluster C_i and C_j. Some FECs are merged after the second stage clustering to eliminate bad mode localization.

Another important aspect of using normalized cut is that, by implementing the recursive n-way cut [18], it is easy to achieve a hierarchical FEC decomposition. By simply adjusting the Ncut value, the number of levels and the number of FECs in each level will be automatically determined, which is ideal for our automatic hierarchical surface abstraction.

3.4 *Shape descriptors for matching*

Once the decomposition is completed, the FECs can be compared for matching purposes. One way to do it is based geometry hashing [6]. In order to facilitate fast matching, we define a feature vector for each FEC as follows:

$$\begin{aligned} Vc = \{ & n_{dev}, gc_{dev}, gc_{mean}, mc_{dev}, \\ & mc_{mean}, gd_{dev}, gd_{mean} \}, \end{aligned} \tag{4}$$

where n_{var} is the deviation of the surface normal, gc_{dev} and mc_{dev} are the deviations of gaussian curvature and mean curvature respectively, gc_{mean} and mc_{mean} are the mean of gaussian curvature and mean curvature respectively, gd_{mean} is the mean geodesic distance of all vertices from the center of the patch normalized by patch size, and gd_{dev} is the corresponding deviation of the geodesic distances. Note that Vc is affine and scale invariant.

4 Surface abstraction based on attributed relational graph

The last step of the shape abstraction procedure is to organize the FECs to form a complete and hierarchical description of the shape. An attributed relational graph is a good choice [19, 20]. Attributed Graph (AG) or Attributed Relational Graph (ARG) is a kind of part-based representation. An ARG is a graph $G = (V, E, A)$, where V is the vertex set, E is the edge set, and A is the attribute set. A consists of unary attribute a_i attaching to each node $v_i \in V$ and binary attribute a_{ij} attaching to each edge $e_{ij} = (v_i, v_j) \in E$. ARG can convert shape matching problems into graph matching problems.

In our framework, each node of the ARG is an FEC, and the associated attribute is the eight-tuple Vc defined in Section 3.4. Adjacent FECs are connected by an edge with e_{ij} being the average geodesic distance between node i and j. The abstraction is achieved by splitting the vertices of the ARG, from the coarsest level to the finest level. This process is illustrated in Figure 2.

The abstraction is extremely useful in tasks such as shape matching and retrieval, since this representation systematically integrate geometric and topological information, which makes it

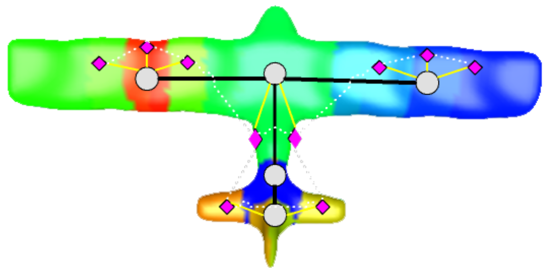

Figure 2: Splitting of vertices to build a hierarchical ARG: circles are graph nodes in the current level, diamonds are graph nodes in the child level. Solid and dashed lines indicate the edges in the current and child level.

easy to handle deformable models and models of high genus. Matching scores from different levels are stored for further queries, ie., surface shapes exhibit different extents of similarity in different levels, which is implied by the abstraction hierarchy. The similarity of two shapes can be measured using the following standards for two different matching purposes:

1. Partial Matching: The highest matching score among different levels of FECs. The level producing the maximum score is called the best matching level.

2. Global Matching: The summation of the scores from all levels.

The following section demonstrates some matching results in order to show the superiority of our method.

5 Experimental results

We have evaluated our algorithm on SHREC dataset. Both rigid model and deformable model are tested. Figure 3 shows the decomposition of an airplane model at different levels.

Figure 4 demonstrates modes found on an airplane model (rigid model) with our adaptive manifold mean shift algorithm. Note that this model is quite noisy, even areas on the wings are quite bumpy. Figure 5 shows the result of FEC-based decomposition and matching, where two bird models are automatically decomposed into FECs and generate the matching correspondence. As for deformable model

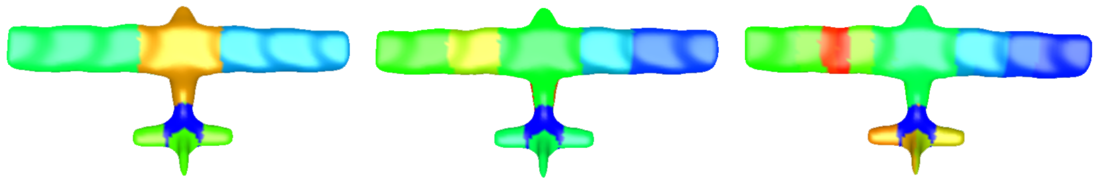

Figure 3: FEC decomposition at 3 different levels

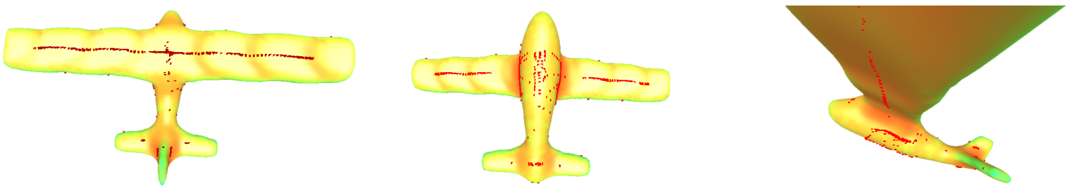

Figure 4: The red points are modes detected by the adaptive manifold mean shift. The model is viewed in three different angles in order to show all the detected modes.

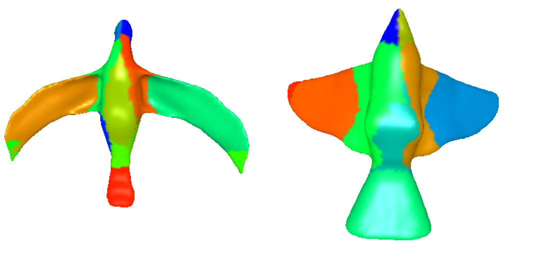

Figure 5: Two bird models matched based on hierarchical FEC decomposition and ARGs. The colors show the decomposition.

matching, we demonstrate the FEC decomposition on two horse models with different poses. They are matched based on their ARG representations. The FEC decomposition and matching correspondence is illustrated in Figure 6. Figure 7 shows that our FEC based graph can handle models with high genus as well, which is very difficult for many algorithms, e.g., the one presented in [14].

We have tested inter- and intra- class matching scores. We selected totally 50 objects from 5 categories. Each category contains 10 objects. We did the inter- and intra-class matching using the hierarchical surface abstraction based on adaptive mean shift and ARGs. The average matching scores between each pair of categories are shown in Figure 8. Based on the matching scores, we can see the framework can perform quite well in shape matching and retrieval.

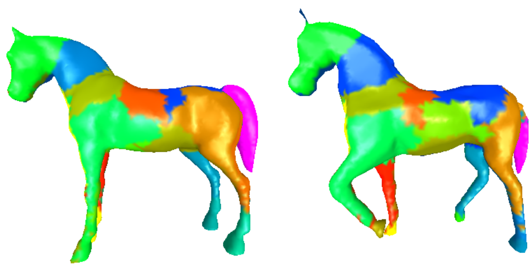

Figure 6: Two horse models with different poses matched with FEC decomposition and ARGs. The FEC decomposition and matching correspondence is illustrated(the matched FECs are colored the same).

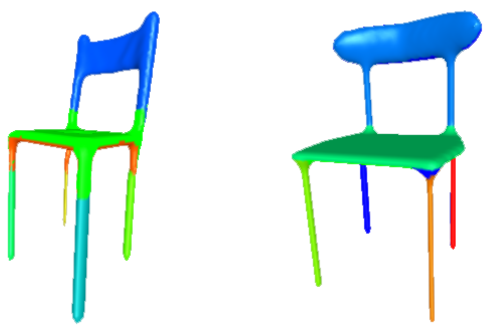

Figure 7: Two objects of genus one matched based on FECs and ARGs.

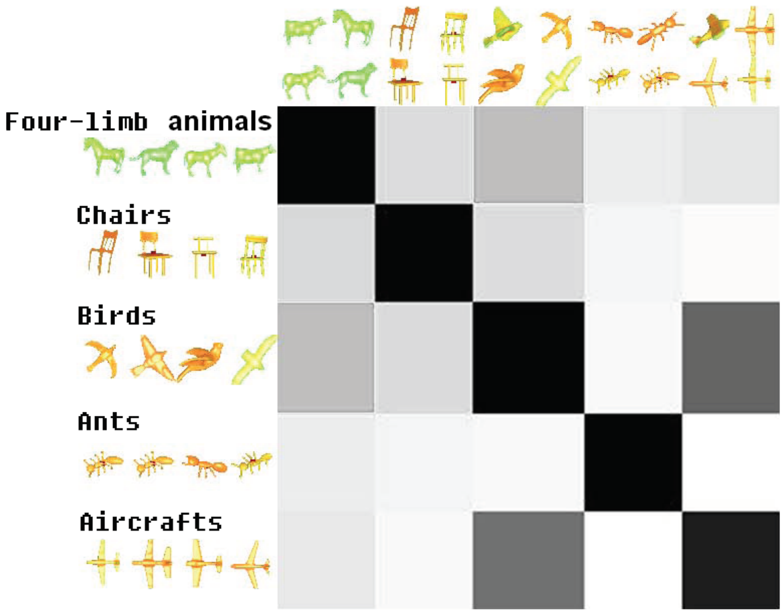

Figure 8: Matching scores placed in a matrix. Each row/column stands for one category, each cell of the matrix indicates the average matching score between two categories. Darker squares stand for higher scores.

6 Conclusion and Discussion

In this paper, we have designed a novel framework for surface abstraction which is well suitable for shape matching and retrieval. Compared to [14], our decomposition is in the composite feature space rather than in the physical space. This gives the possibility to make the components geometric and topological feature enriched instead of only topologically meaningful. In addition, our method can handle objects with high-genus topologies, and is insensitive to resolution and noise. Furthermore, our ARG is built in a hierarchical way towards a full abstraction of the input surface. The surface objects can be compared at different levels for fast computation. The framework can be adapted to different applications.

However, due to the lack of extra knowledge, the feature space is built solely from local geometry such as curvature and normal. This limitation can be exceeded by introducing more features, such as texture and color. As the dimension of the feature space increases, the well known curse of dimension need to be treated properly then. This is a very good direction to explore in the near future. At present, the algorithms work fine in the seven dimensional feature space. In the future, we will also spend more effort on applying the framework on motion analysis and animation synthesis.

Acknowledgements

This work is supported in part by the following research grants award to Dr. J. Hua: the National Science Foundation grants IIS-0713315 and CNS-0751045, the National Institute of Health grants 1R01NS058802-01A2 and 2R01NS041922-05A1, the Michigan Technology Tri-Corridor grants MTTC05-135/GR686 and MTTC05-154/GR705, and the Michigan 21st Century Jobs Funds 06-1-P1-0193.

References

[1] Curless. B. Allen and Z. Popovic B. Exploring the space of human body shapes: Data-driven synthesis under anthropometric control. In *Proceedings of Conference*

on Digital Human Modeling for Design and Engineering, 2004.

[2] Hongbin Wang and Hua Li. Cartoon motion capture by shape matching. In *Proceedings of 10th Pacific Conference on Computer Graphics and Applications*, pages 454–456, 2002.

[3] Johan W.H. Tangelder and Remco C. Veltkamp. A survey of content based 3D shape retrieval methods. In *Proceedings of the Shape Modeling International*, volume 0, pages 145–156, 2004.

[4] S. E. Palmer. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9:441–474, 1977.

[5] E. Wachsmuth, M. W. Oram, and D. I. Perrett. Recognition of objects and their component parts: reponses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, 4:509–522, 1994.

[6] Ran Gal and Daniel Cohen-Or. Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics, January*, 25(1), 2006.

[7] Philip shilane and Thomas Funkhouser. Distinctive regions of 3D surfaces. *ACM Transactions on Graphics*, 2007.

[8] Yushen Liu, Min Liu, Daisuke Kihara, and Karthik Ramani. Salient critical points for meshes. In *Proceedings of the ACM symposium on Solid and physical modeling*, pages 277–282, 2007.

[9] R. Wessel, M. Novotni, and R. Klein. Correspondences between salient points on 3D shapes. In *proceedings of Vision, Modeling, and Visualization*, pages 365–372, 2006.

[10] Chang Ha Lee, Amitabh Varshney, and David W. Jacobs. Mesh saliency. In *SIGGRAPH*, 2005.

[11] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson. Skeleton based shape matching and retrieval. In *SMI*, pages 130–139, 2003.

[12] N. Lyer, Y. Kalyanaraman, K.Lou, S. Janyanti, and K. Ramani. A reconfigurable 3D engineering shape search system parti: shape representation. In *DETC*, 2003.

[13] N. Lyer, Y. Kalyanaraman, K.Lou, S. Janyanti, and K. Ramani. A reconfigurable 3D engineering shape search system partii: database indexing, retrieval and clustering. In *DETC*, 2003.

[14] Gary K.L. Tam and Rynson W.H. Lau. Deformable model retrieval based on topological and geometric signatures. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):470–482, 2007.

[15] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

[16] Yizong Cheng. Mean shift, mode seeking, and clustering. *TPAMI*, 17(8):790–799, 1995.

[17] Ariel Shamir, Lior Shapira, and Daniel Cohen-Or. Mesh analysis using geodesic mean-shift. *The Visual Computer*, 22(2):99–108, 2006.

[18] Jianbo Shi and Jitendra Malik. Normalized cut and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 2000.

[19] Dong-Qing Zhang and Shih-Fu Chang. Detecting image near-duplicate by stochastic attributed relational graph matching with learning. In *ACM Multimedia*, pages 877–884, 2004.

[20] Dongqing Zhang. Statistical part-based models: Theory and applications in image similarity, object detection and region labeling. *PhD Thesis Graduate School of Arts and Sciences, Columbia University*, 2005.