2014 International Conference on Virtual Reality and Visualization

View Selection of 3D Objects Based On Saliency Segmentation

Honglei Han[123], Jing Li[2], Wencheng Wang[2], Huiwen Zhao[1], Miao Hua[23]
[1] Animation School, Communication University of China, Beijing, PRC
[2] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, PRC
[3] University of Chinese Academy of Sciences, Beijing, PRC
E-mail: hanhonglei@sina.com

***Abstract*—Getting good viewpoints has been considered important for promoting the efficiency when investigating a model. Many view selection methods therefore have been proposed. In particular, measuring semantic meaning of the model features through segmentation is regarded more effective to get optimal viewpoints. Unfortunately, the semantic meanings of the model are usually obtained through experiences, which are not easy to apply into practices. Given these considerations, a new method is proposed for view selection via saliency segmentation in this paper. Due to the fact that mesh saliency computation can recognize the features of a model in a way that similar to human perception and also a novel viewpoint ranking equation has been developed, our method is more reasonable than existing methods. The results of experiments show the effectiveness of our method, and it is consistent with the results of a user study we conducted.**

Keywords-mesh saliency; view selection; segmentation

I. INTRODUCTION

3D models have been widely used in many fields such as computer game, virtual reality, film production *et al.* And various modeling software help 3D artists to create a large number of 3D models. With the rapid increase of digital models, it is urgent to find high efficient methods to index and browse models. However, the usual way of manipulating 3D models with 2D interface makes it hard for users without specialized knowledge to effectively observe 3D models. To this, automatically create good views for users would have great help.

There is an agreement that the best view is closely related to its salient components. For example, the eye is a salient feature of a head model, so it should be visible in the image rendered by the best viewpoint. Nonetheless, the majority of prior work on this topic deals with just low level geometric properties of the object surfaces. So, the aforementioned eye feature is not easy to be identified using these geometric based methods.

Recent years, there have been view selection methods based on semantics. Most of these methods need to compute the segmentation on object surface. This is mainly due to the fact that the structure information of 3D object can be extracted by carefully designed model segmentation method. It is convenient to reflect high-level features based on the semantic information. However, the semantic importance for each segment is hard to be automatically calculated. In other words, it is difficult to get reasonable view selection results through segmentation which is lack of appropriate semantic importance weights.

Inspired by these semantic based method, we argue that the best view should be evaluated taking into consideration the different level of saliency components, or features, of a 3D shape. That is, the quality of a view should be related to the visual importance of the displayed features. In particular, we propose a new scoring function for selecting the optimal viewpoint which integrates visibility criteria with saliency weighted features which are visible in a given direction. Mesh saliency can reflect the visual perception of model features. It is particularly appropriate to serve as perception importance metric. Therefore, in this paper we cluster adjacent faces with similar saliency into one patch and the visual importance of each patch is determined by saliency of faces belong to it. This saliency-based segmentation can truly reflect each patch's visual importance and has advantages of possessing structural information of 3D objects.

The paper is structured as follows: First, we briefly outline the most related work in viewpoint ranking and high level information extracting of 3D geometry in Section 2. Then, we describe how to make mesh segmentation efficiently by mesh saliency in Section 3. In Section 4, we introduce how to do view selection according to the saliency-based segmentation. In Section 5, we report and analyze the results of comparison experiments and user studies to verify the advantages our method. Finally, the final conclusions have been drawn in Section 6.

II. RELATED WORK

A. View selection methods

View selection has attracted great attention in recent years. Secord *et al.* [1] reviewed these methods in detail. They classified these methods into five types according to the attributes of views used in measuring view goodness: (1) area, *e.g.* projected area, surface visibility [2] and viewpoint entropy [3, 4]; (2) contour [5, 6]; (3) depth [7, 8]; (4) surface curvature [9-13]; (5) semantic [7, 14-16].

Methods of the first four types mainly use elementary geometrical information to rank viewpoints. Although some of them introduced information theory (such as Shannon Entropy) to quantify information captured from 3D model [3, 4, 11, 12], lacking structural information makes it hard to truly reflect the perceptual habits of human beings[15]. As the 5th type evaluates viewpoints through high-level even semantic information extracted from 3D object, it avoids the

978-1-4799-6854-1/14 $31.00 © 2014 IEEE
DOI 10.1109/ICVRV.2014.12

problem of other four types [15, 16]. That is because the mesh segmentation represents the high-level semantic meaning of the model. However, reasonable metrics to evaluate the semantic importance for each segment are lacking in these methods. Therefore, they need to manually decide the visual importance of each segment, which makes it sensitive to the assignment bias and does not guarantee getting reasonable results.

Our method proposed in this paper belongs to the fifth type. However, instead of manually setting semantic importance value, we employed mesh saliency as mesh segmentation metric to automatically generate perceptual importance weighted segmentation results.

B. Mesh segmentation algorithm

The question of how to extract high-level structure information from geometry elements plays a fundamental role in our method. According to [17], shape understanding and semantic-based object representation must rely on feature extraction and structure extraction from 3D meshes. As an important way to extract structural information, many 3D mesh segmentation algorithms have been proposed over last decade and several papers have made a thorough survey of these techniques [17-19].

Some mesh segmentation methods are fully automatic: they either extract regions around contours [20] or, like ours, cluster similar mesh elements into regions [21]. Both of these approaches concentrate on analyzing pure geometric or topological information of the input mesh, such as, shape diameter function [21], planarity and normal directions [22]. Unlike the pure geometrical or topological criterion as mentioned above, mesh saliency is used as criterion to get segmentation of mesh in our study. As mesh saliency can model human eye movements significantly [23], taking mesh saliency as the criterion is much easier to decompose mesh into patches with different level of perceptual importance. In this paper, the graph-based segmentation method in image processing [24] is introduced to handle 3D mesh segmentation.

III. MESH SEGMENTATION BY MESH SALIENCY

In 1987, Biederman proposed "Recognition by Component Theory" in the literature [25]. The theory realizes that complex objects can be efficiently recognized by decomposing their structures into simple components. Based on this theory, some mesh segmentation based view selection methods are more likely to get reliable results [12, 15]. However, the segmentation process in these methods are based on low level geometry information such as folding angle or curvature, and importance weights of each component are assigned as empirical values [15] which are difficult to meet the real human recognition metric. In view of this, we treat mesh saliency [9] which are proved to be reasonable when modelling human eye movements [23] as the metric to do mesh segmentation. These saliency featured patches will be treated as basic units when computing viewpoint quality.

Our view selection method depends on the ability of saliency–based mesh segmentation to reflect perceptual-dependent content of 3D objects. Therefore, in the computation of segmentation we take two main measures to improve this ability:

1. Simplify model to get a non-redundant mesh.
2. Cluster vertex saliency of simplified mesh into patches with similar saliency.

In the following sections, we will discuss them in detail.

A. Efficient mesh saliency computation

We observe that in the process of recognizing a complex model, many high frequency geometry details cannot be noticed by human eyes at a certain distance, which has little contribution to the model recognition. Moreover, as saliency computation is usually closely related to the complexity of model, such details waste too much computation resource. Such unnoticeable details can be removed by a sophisticated model simplification process. The mesh saliency and the following view selection computation will be greatly improved using the simplified mesh.

To guarantee visual perception on model not be changed in the simplification, we need to find the maximum length in model space that would not be perceived by human eyes, denoted by E_m, and use this parameter to control the simplification. We stop the process of simplification as soon as the change caused by simplification exceeds such parameter. Here, we use Contrast Sensitivity Function (CSF) [26] to calculate E_m.

In the process of mesh simplification, we adopt Quadric Error Metrics (QEM) method [27]. It selects an edge with minimum cost, then merges two vertices of this edge, and finally adjusts merged vertex to an optimal position. Such collapse operation continues until the collapse error exceeds a threshold. Here, the collapse error is defined as the sum of squared distances from new generated vertex to its associated planes. To guide the simplification by visual perception, we set the threshold as E_m^2, which means the original mesh can be simplified to an optimal level (we call it non-redundant version) that the removed details are not able to be noticed.

After mesh simplification, we adopt the method used by [9, 23] to calculate vertex saliency for non-redundant mesh. It uses a center-surround mechanism that is inspired by the human visual system. The reader should refer to [9]. In our experiments, the calculation efficiency of mesh saliency using non-redundant mesh is enhanced 6 to 7 times than using the original one.

B. Graph based segmentation

In order to obtain a visual importance related decomposition of the mesh, we employ a mesh segmentation method using the metric of mesh saliency[9] in a way similar to the graph-based segmentation method [24], which has been successfully used in the field of 2D image processing. Such a segmentation method has two properties that are important for our view selection. One is its ability to capture perceptually important groupings or regions. The other is its high efficiency.

We make the following modifications to guarantee this 2D image segmentation method can be used to handle the 3D mesh:

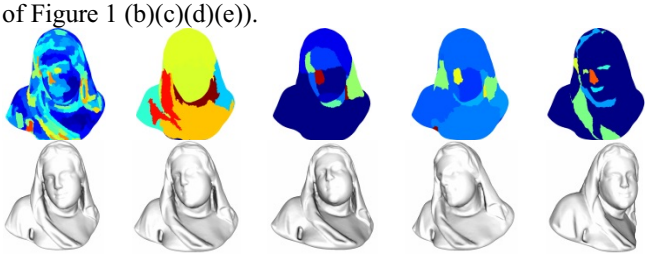
1. Pixel neighbor relationship is replaced with adjacent relationship between faces in the 3D mesh.

2. Difference between pixels is replaced with absolute value of saliency difference between adjacent faces

3. The saliency of every face is the averaged saliency of its three vertexes.

In the implementation, we set the threshold of saliency difference between adjacent patches to 0.4% of the difference between the largest and the smallest value of vertex saliency. And the number of faces in the smallest patch is set to 0.05% of the total number of faces in the mesh to avoid too small patches.

From the comparisons in Figure 1, it is easy to see that saliency is more appropriate to server as perception importance metric than curvature and other geometry metric. Images in the first row of Figure 1(b)(c)(d)(e) are segmentation results using methods similar as [15], which firstly segment mesh using classic segmentation methods based on geometry metrics, and then calculate every segments' visual importance based on its member faces' saliency (warmer color indicates higher importance). While in [15], they just assign each type of patch different empirical importance values in second step. Some of important feature are lost in these geometry based segmentation results (such as eyes, mouth, and the hole near the armpit in the first row of Figure 1 (b)(c)(d), and hair in the first row of Figure 1 (e)). The decomposition result using our method captures most of important features (see the first row of Figure 1(a)). So the best viewpoint selected by our method (the second row of Figure 1(a)) is much closer to canonical view [7] compare with the others (the second row of Figure 1 (b)(c)(d)(e)).



(a) Ours (b) FitPrim [22] (c) RandCuts [20] (d) ShapeDiam [21] (e) Benchmark [19]

Figure 1. Mesh decomposition (the first row) and their best view selection results (the second row) produced using different segmentation metric

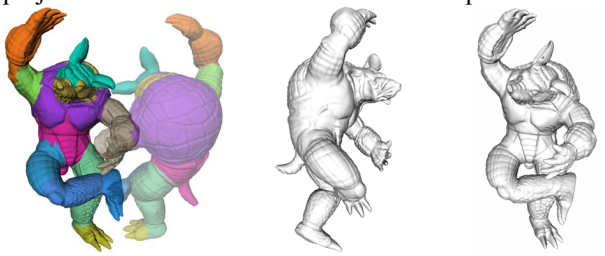
IV. SCORING VIEWPOINTS BY SALIENCY SEGMENTATION

As saliency-based mesh segmentation can truly reflect perceptual related model features, we can then make use of it to rank viewpoints effectively. In the following, we first discuss the main problem exist in recent viewpoint ranking approach by mesh segmentation, and then discuss how to use saliency segmentation as viewpoint scoring unit.

Mortara *et al.* introduce a view scoring formula taking semantic segmentation as computation unit [15]. They considered that meaningful segmentation contains more semantic information. Formula of scoring viewpoint w in their method is expressed as:

$$Score(w) = N_{vf}(w) \cdot \sum_c V(c,w) R(c) W(c) \quad (1)$$

, where N is the total number of patches which are visible from current viewpoint, c is a visible segment, and R is the area ratio of this patch to the whole mesh, W is a weight assigned based on the type of patch. V is the degree of visibility, which is calculated by the sum of the ratio between projected area and surface area of faces in a patch.



(a) Segment results (b) Best view in [15] (c) Our best view

Figure 2. Armadillo segmentation and view selection result of [15] and ours

However, the main problem of using equation 1 in [15] is the weight W for each patch is manually assigned according to algorithms used in segmentation and the type of patches. As one can see in Figure 2, Armadillo model is segmented into 12 patches using FP method [22] (Figure 2a), all the fitting type of each patch is cylinder, sphere or cone, their importance weight are equal according to [15]. The importance weight of each patch becomes meaningless, so the best viewpoint got by their method (Figure 2(b)) is not the same as what people expected compared with the result of our method (Figure 2(c)) which is more close to canonical view[7].

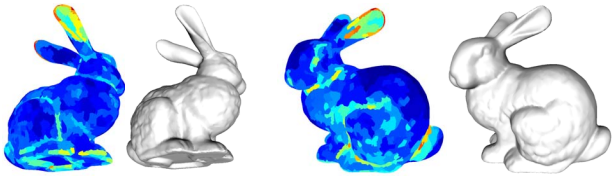
To solve this problem, we use saliency-based segmentation results produced in previous section as viewpoint evaluation data:

$$Score(w) = N_{vf}(w) \cdot \sum_c V(c,w) R(c) S(c) \quad (2)$$

, where S is the saliency value of patch c in saliency segmentation. We normalize S to the interval of $[\delta, 1+\delta]$, where δ is a positive value far less than 1.This ensures the patch with the lowest saliency still has a certain influence.

However, it is not appropriate to take the patch area as weight directly in equation 2. For instance, when observing a person, though the area of eyes is smaller than that of body, eyes have far more semantic importance. So we prefer put our focus on its face. It is also not reasonable to allow the number of visible patches to appear in the formula directly as parameter, as counting up of visible patches' importance has denoted this factor. So we revised equation 2 as follows:

$$Score(w) = \sum_c V(c,w) S(c) \quad (3)$$



(a) Equation 2 (b) Equation 3

Figure 3. The best viewpoints using different scoring equations

Preferred viewpoint obtained by equation 2 is easy to be affected by a few large patches with high saliency value which is alleviated in equation 3. It can be seen from Figure 3, the best viewpoint selected by equation 2 is from bunny's back (Figure 3(a)) which is inappropriate compare with canonical view selected by equation 3(Figure 3(b)).

V. EXPERIMENT RESULTS AND ANALYSIS

We made several experiments on a PC with Intel Core i7-860 CPU (2.80GHz, 8CPUs), 12GB RAM and a NVIDIA GeForce GTX 460 GPU (1GB RAM). In order to evaluate superiorities of our method, we compare results produced by our method and other state-of-arts methods, and conduct a series of preliminary user studies.

For each test model, we create a view sphere, and then 258 candidate viewpoints are equally distributed on this sphere. We calculated and sorted the scores of all these viewpoints by equation 3 using saliency segmentation we proposed. The highest quality viewpoint will be selected as preferred view of the model.

A. Comparison with other methods

We have compared our method with the semantic-driven approach introduced in [15]. As can be seen from Figure 1 and Figure 2, best views selected by our method are more intuitive and informative. This is mainly due to two facts: one is we take saliency segmentation, which possesses perceptual related weights, as the basic elements of view selection; the other is we revised the viewpoint scoring formula used in [15].

We also compared our results with two other recent state-of-arts works by Leifman *et al.*[10] and Secord *et al.*[1].

Figure 4 shows the results compared with [10]. Taking symmetry into consideration, best views of Dinosaur and Motorcycle by the two methods are almost the same. But for Lion model, our method can generate more satisfied canonical view. Viewers can capture informative characteristics more easily by the best view of our method, *e.g.* length of lion's body and tail.

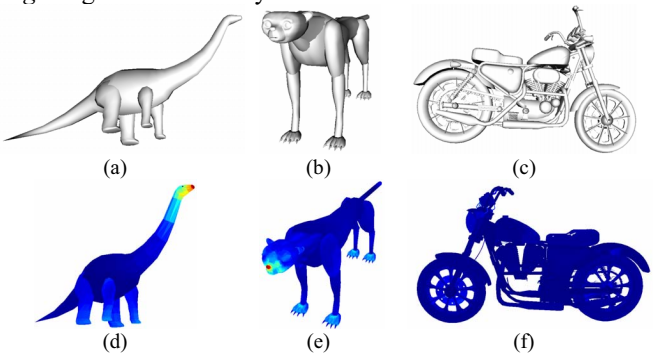


Figure 4. Best viewpoints selected by [10] (a, b, c) and by our method (d, e, f)

Another state-of-arts method [1] selected 20 pairs of views for each of the 16 models (3,840 images in total). Users were asked to select their preferred view from two different compared views. After analyzing 30-40 valid questionnaires, they got the best viewpoint for each model that is accepted by most people (Figure 5 illustrates some comparing results between this method and ours). We use the forced choice pair-wise comparison method to evaluate preference of view selection results produced by [1] and our method. 16 pairs of best viewpoint for each model are generated by [1] and our method. Then, they are reviewed by 34 volunteers. The statistical results can be seen in Figure 5. Considering the symmetry, a large portion of the best viewpoints obtained by our method are very close to ones in [1], so their preference percentages are similar (See Figure 5(a)(b)(c)(d)). However, some view selection results are obviously different between two methods (See Figure 5(e)(f)(g)(h)). The best viewpoints in [1] for Ant, Eagle and Chair (Figure 5(e)(f)(g)) are placed in back that is unnatural when observing objects, so their preference percentages are lower than ours. But for Octopus model, our method generated inappropriate view, as many visual important features distributed in the bottom of the object. More importantly, getting the best viewpoints in [1] needs a large number of user interactions, our method is fully automatic.

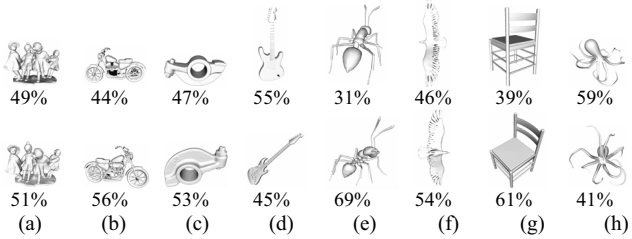


Figure 5. Some results obtained by state-of-arts method in [1] (the 1st row) and by our method (the 2nd row). The numbers under each image represents preferred percentages in our user study results

B. User studies

We conducted a preliminary user study to verify the discrimination of our method. Based on the classification method in [28], we tried to select models of various kinds as test dataset. We selected 28 models of different kinds finally. We used our method to rank all candidate viewpoints for all test models. The viewpoint with the highest score using our method was assigned a score 5 and that with the lowest score is assigned a score 0. Then 5 viewpoints were selected whose scores are 5, 3.75, 2.5, 1.25 and 0 to represent the best, acceptable, medium, disappointed and the worst viewpoints. After that, users were asked to re-order these 5 viewpoints. We used inverse pair to measure consistency between sorting results by users and by our method. Specifically, let $(A(1), \ldots, A(k))$ be a sequence of k distinct numbers. If $i < j$ and $A(i) > A(j)$, then the pair (i, j) is called an inversion of A. The fewer the inversion is, the more consistent with user sorting results and our method. We collected 29 valid questionnaires from the candidates. Figure 6 shows numbers of inversions for all test models. It is easy to see that numbers of inversions for most models are less than 1. Only three models have more than 2 on average. They are Fighter, Brain and Bunny. This may be related to characteristics of these three models: quality differences of viewpoints from different observing directions are small.

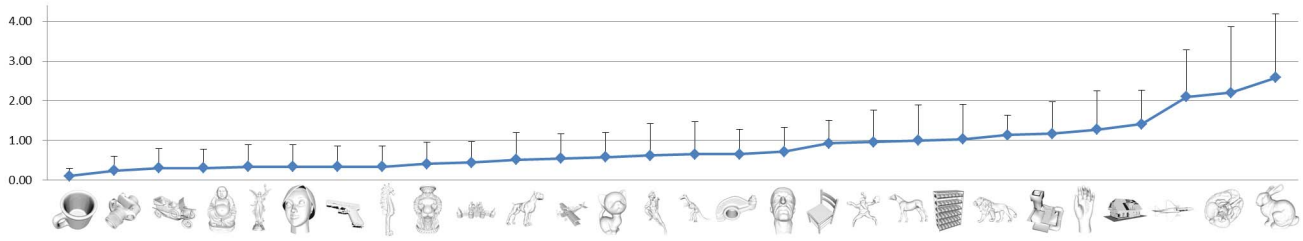


Figure 6. The numbers of inversion and average deviations for 28 test models. The thumbnails below the chart are the best viewpoints of test models obtained by our method.

VI. CONCLUSIONS AND LIMITATIONS

In order to rank viewpoints of 3D object more consistent to human habits, we proposed a new mesh decomposition method based on mesh saliency. It combines mesh saliency and high-level global structural information. The viewpoints which allow seeing more important patches are preferred. Experiments show that view selection results of our method is more consistent with perceptual habit of human eyes, benefiting from the inherent perceptual property of saliency segmentation.

According to the description in Section 3 and 4, both of the processes of mesh segmentation and view selection are dependent on the calculation results of mesh saliency. If mesh saliency fails to be consistent with the perception of human eyes, we will get improper results (Figure 7). Maybe some other perception dependent parameters (*e.g.* surface regions of interest introduced in [10]) should be involved as criteria of segmentation meshes.

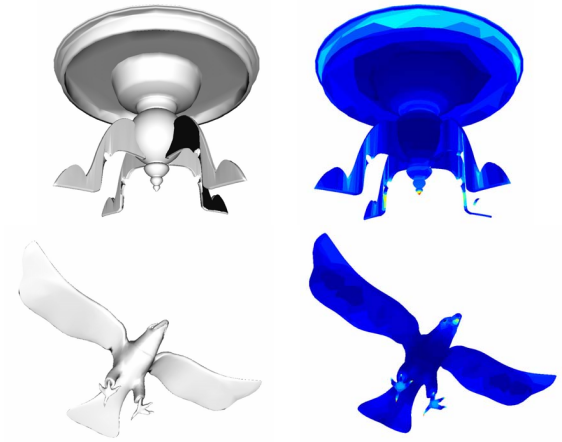


Figure 7. Two failure examples and their corresponding saliency segmentations

ACKNOWLEDGMENTS

This work is supported by the National Key Technology R&D Program under Grant No. 2012BAH62F03, and the National Natural Science Foundation of China under Grant No. 60873182, 61379087. All the test models in this paper are scratched from open model databases on the internet including AIM@SHAPE Shape Repository, Stanford University, and Princeton Shape Benchmark [28]. We want to express our appreciation to the authors of these 3D models.

REFERENCES

[1] Secord A, Lu J, Finkelstein A, Singh M and Nealen A. Perceptual models of viewpoint preference. ACM Trans. Graph., 2011, 30(5): 1-12.

[2] Plemenos D and Benayada M. Intelligent display in scene modeling. new techniques to automatically compute good views. In International Conference GraphiCon, pp. 1-5.

[3] Vázquez P-P, Feixas M, Sbert M and Heidrich W. Automatic View Selection Using Viewpoint Entropy and its Application to Image-Based Modelling. Computer Graphics Forum, 2003, 22(4): 689-700.

[4] Serin E, Sumengen S and Balcisoy S. Representational image generation for 3D objects. The Visual Computer, 2013, 29(6-8): 675-684.

[5] Feldman J and Singh M. Information along contours and object boundaries. Psychological Review, 2005, 112(1): 10.

[6] Vieira T, Bordignon A, Peixoto A, Tavares G, Lopes H, Velho L and Lewiner T. Learning good views through intelligent galleries. Computer Graphics Forum, 2009, 28(2): 717-726.

[7] Blanzfil V and Btilthoff H H. What object attributes determine canonical views? Perception, 1999, 28: 575-599.

[8] Stoev S L and Strasser W. A case study on automatic camera placement and motion for visualizing historical data. In Visualization, 2002. VIS 2002. IEEE, 1-1 Nov. 2002, pp. 545-548.

[9] Lee C H, Varshney A and Jacobs D W. Mesh saliency. ACM Trans. Graph., 2005, 24(3): 659-666.

[10] Leifman G, Shtrom E and Tal A. Surface regions of interest for viewpoint selection. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 16-21 June 2012, pp. 414-421.

[11] Page D L, Koschan A F, Sukumar S R, Roui-Abidi B and Abidi M A. Shape analysis algorithm based on information theory. In Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, 14-17 Sept. 2003, pp. I-229-232 vol.221.

[12] Polonsky O, Patané G, Biasotti S, Gotsman C and Spagnuolo M. What's in an image? The Visual Computer, 2005, 21(8): 840-847.

[13] Yang Y-L and Shen C-H. Multi-Scale Salient Features for Analyzing 3D Shapes. Journal of Computer Science and Technology, 2012, 27(6): 1092-1099.

[14] Fu H, Cohen-Or D, Dror G and Sheffer A. Upright orientation of man-made objects. In ACM SIGGRAPH 2008 papers, Los Angeles, California, pp. 1-7.

[15] Mortara M and Spagnuolo M. Semantics-driven best view of 3D shapes. Computers & Graphics, 2009, 33(3): 280-290.

[16] Laga H. Semantics-driven approach for automatic selection of best views of 3D shapes. In Proceedings of the 3rd Eurographics conference on 3D Object Retrieval, Aire-la-Ville, Switzerland, pp. 15-22.

[17] Shamir A. A survey on Mesh Segmentation Techniques. Computer Graphics Forum, 2008, 27(6): 1539-1556.

[18] Attene M, Katz S, Mortara M, Patane G, Spagnuolo M and Tal A. Mesh Segmentation - A Comparative Study. In Shape Modeling and Applications, 2006. SMI 2006. IEEE International Conference on, 14-16 June 2006, pp. 7-7.

[19] Chen X, Golovinskiy A and Funkhouser T. A benchmark for 3D mesh segmentation. ACM Trans. Graph., 2009, 28(3): 1-12.

[20] Golovinskiy A and Funkhouser T. Randomized cuts for 3D mesh analysis. ACM Trans. Graph., 2008, 27(5): 1-12.

[21] Shapira L, Shamir A and Cohen-Or D. Consistent mesh partitioning and skeletonisation using the shape diameter function. The Visual Computer, 2008, 24(4): 249-259.

[22] Attene M, Falcidieno B and Spagnuolo M. Hierarchical mesh segmentation based on fitting primitives. The Visual Computer, 2006, 22(3): 181-193.

[23] Kim Y, Varshney A, Jacobs D W and Guimbretière F. Mesh saliency and human eye fixations. ACM Trans. Appl. Percept., 2010, 7(2): 1-13.

[24] Felzenszwalb P and Huttenlocher D. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 2004, 59(2): 167-181.

[25] Biederman I. Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review, 1917, 1000(2): 115-147.

[26] Kelly D H. Motion and vision. II. Stabilized spatio-temporal threshold surface. J. Opt. Soc. Am., 1979, 69(10): 1340-1349.

[27] Garland M and Heckbert P S. Surface simplification using quadric error metrics. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pp. 209-216.

[28] Shilane P, Min P, Kazhdan M and Funkhouser T. The Princeton Shape Benchmark. In Shape Modeling Applications, 2004. Proceedings, 7-9 June 2004, pp. 167-178.