Chapter 64
A Novel Quantitative Evaluation Metric of 3D Mesh Segmentation

Xiao-peng Sun, Lu Wang, Xingyue Wang and Xiaona Zhao

Abstract This paper presents a novel quantitative metric for comparison of 3D mesh segmentations, Ultimate Measurement Accuracy, basing on the screening data sets, to support our quantitative comparisons of seven recently published mesh segmentation algorithms; and the experiment results suggest that, our metrics is robust to degenerative segmentation and hierarchical refinement, stable to the imprecision of cut boundaries.

Keywords 3D segmentation evaluation · Quantitative evaluation · Ultimate measurement accuracy · Misclassified error

64.1 Introduction

3D mesh segmentation benefits various algorithms in Digital Geometry Processing, such as modeling, simplification, texture synthesis, collision detection, skeleton extraction, compression, and 3D shape retrieval and the establishment of the corresponding relations in morphing and deformation [1].

In the last several years, a number of segmentation algorithms have been proposed to cut 3D mesh into simple sub-meshes, including a few components with higher level shape semantic, or lower level disk-like patches, and many of them aim at a meaningful segmentation, but little work on the qualitative evaluation of quality, even little on the quantitative evaluation. [2–4].

The main contributions of this work is that, we investigate a novel evaluation metrics for quantitative comparing mesh segmentations—UMA, and discuss its properties; To improve the accuracy of our three metrics, we establish a subset of 3D mesh segmentation Benchmark as our experimental dataset.

X. Sun (✉) · L. Wang · X. Wang · X. Zhao
Computer Systems Institute, Liaoning Normal University, Dalian 116029, China
e-mail: cadcg2008@gmail.com

64.2 Related Work

Most work evaluated their segmentation only by the vision; little work has been done on the qualitative and quantitative quail evaluation of segmentation.

The qualitative criteria frequently used to evaluate the automatic segmentation algorithms in a qualitative way: exhibit the segmentation image of the same set of meshes side by side, and point out which results look good or not. And the criteria include: the robustness and complexity, the number of control parameters, the hierarchical and multi scale properties, whether to produce over-segmentation, the smoothness of the segmentation boundary, and whether the segmentation results have a clear visual geometry semantics (meaningful), whether the segments boundaries lie on concave seams and valleys, as predicted by the minima rule.

Quantitative evaluation of 3D segmentation first appeared in 2009. To address the key problem of quantitative evaluation, Benhabiles et al. [2] proposed a ground truth corpus: two objective dissimilarity measures, which provide a quantitative comparison between the segmentation algorithms. The limitations of this work are the number of model and category, and the meaningful of manual segmentations.

Comparing with the work of Benhabiles et al. [2], this data set of Chen et al. [3] has a greater size and a higher sampled distribution. To provide a basis for quantitative comparisons between the human-generated segmentations and automatic segmentation [5–10] for the same mesh, they propose four quantitative evaluation metrics. In 2010, Benhabiles et al. [4] present an extensive experimental comparison of existing similarity metrics addressing the quality assessment problem of mesh segmentation and introduce a new metric, and basing on their own corpus and Chen et al. corpus [3], they get a better performs, but their work has a drawback with the accuracy of probability of two vertices belonging to the same segment in the ground-truth set.

All of these works provide a benchmark basing on a ground-truth corpus of human segmented 3D meshes. Using a set of similarity metrics, they evaluate the segmentation algorithm by quantifying the consistency between the reference segmentations of the ground-truth corpus and those obtained by this algorithm on the same models.

Some problems with the benchmark as follow: First, the manual segmentation results produced by different volunteers show significant subjectivity and randomness. Second, although Consistency Error evaluation metric proposed independently by Chen et al. and Benhabiles et al. are similar, they produce a significant different evaluation metrics for the same segmentation algorithm. Third, the four kinds of evaluation criteria proposed by Chen et al. still produce significant different evaluation metrics for the same segmentation algorithm with others. Finally, all the six quantitative metrics produce very differences evaluation to which by human visual perception.

In this paper, we first analyze the serious inconsistencies, errors and other problems of the data set produced by manual interaction segmentation in [2–4], and

build a subset of Chen's dataset with higher consistency and accurate. Finally, we perform quantitative comparison of seven automatic mesh segmentation algorithms according to the new metric UMA, and our other two metrics, FME and AME [11].

64.3 Ultimate Measurement Accuracy

The ultimate goal of 3D mesh segmentation is to obtain the precise measurement of target feature, so we can evaluate the quality of segmentation by means of measuring and calculating the target feature. The common features frequently-used to measure the target feature precisely include target area, perimeter, circular, eccentricity, shape parameters, Ultimate Measurement Accuracy (UMA) et al.

64.3.1 Shape Feature Form Factor

First, we define the sub-grid shape feature Form Factor (FF) as follows: Let S be a 3D boundary mesh, and S1 the set of sub-grid generated by manual interactive segmentation, i.e., $S1 = \{S_1^1, S_1^2, \ldots, S_1^m\}$, and S2 the automatic segmentation produce by the given algorithm, i.e., $S2 = \{S_2^1, S_2^2, \ldots, S_2^n\}$, m and n are the number of segments. Here we require m equals to n. For each sub-grid S_1^i in S1, note its match position in S2 is sub-grid S_2^j, i.e., there is a correspondence $j = p(i)$. We define the Form Factor of arbitrary sub-grid S_1^i in S1 as follows:

$$FF(S_1^i) = \sum_{a=1}^{N_e} L(S_1^i, e_a) \Big/ \sum_{b=1}^{N_f} A(S_1^i, f_b) \tag{64.1}$$

where e_a and f_b are the arbitrary boundary edge and face in sub-grid S_1^i respectively, and $L(S_1^i, e_a)$ is the length of edge e_a, $A(S_1^i, f_b)$ is the area of f_b, N_e and N_f are the number of boundary edges and faces of S_1^i respectively. Similarly, the Form Factor of arbitrary sub-grid S_2^j in S_2 is defined as follows:

$$FF(S_2^j) = \sum_{a=1}^{N_e} L(S_2^j, e_a) \Big/ \sum_{b=1}^{N_f} A(S_2^j, f_b) \tag{64.2}$$

where the e_a, f_b, L, A, N_e and N_f are defined similarly with that as before. Then, we obtain the Form Factors of all the sub-grid produced by manual interactive segmentation and automatic algorithm segmentation.

64.3.2 Ultimate Measurement Accuracy

With the manual interactive segmentation S_1 as criterion, the UMA of the segmentation S_2 produce by the given algorithm is defined as follows:

$$UMA(S_1, S_2) = \frac{1}{m} \sum_{i=1, j=p(i)}^{m} \frac{FF(S_1^i) - FF(S_2^j)}{\left\| S_1^i \right\|}$$ (64.3)

where $\left\| S_1^i \right\|$ is the area of S_1^i. The smaller the value of $UMA(S_1, S_2)$, the closer the automatic algorithm segmentation results with manual interactive segmentation results, the better the performance of segmentation algorithm, or vice versa.

To a given automatic segmentation algorithm, for each 3D boundary-mesh S in the data set of benchmark, one of its manual interactive segmentation is S_1^i and its segmentation produced by the given automatic algorithm is $S_2 = \left\{ S_2^1, S_2^2, \ldots, S_2^n \right\}$, note the Ultimate Measurement Accuracy as $UMA(S_1^i, S_2)$, then for the set of manual interactive segmentation $\{S_i\}$, the $UMAs$ of this automatic segmentation algorithm is defined as $\sum_i UMA_S(S_1^i, S_2)/n$, where n is the number of manual interactive segmentation models.

For the manual segmentation results of a data set with N different meshes, we define its quantitative evaluation metrics using Ultimate Measurement Accuracy as $UMA = \sum_j UMAS/N$. The smaller the UMA is, the closer the automatic algorithm segmentation results with manual interactive segmentation results, the better the performance of segmentation algorithm.

64.4 Results and Analysis

64.4.1 Experimental Environment

Our experiments are performed on a subset of the benchmark data set proposed by Chen et al. [3], but reject 3 kinds degenerative segmentation results as following: First, the segmentation with the sub-grid number less than or equal to 3; Second, and the ratio of that the face number in sub-grid to the total face number of the original mesh less than or equal to 0.01; Third, the segmentation results without a clear meaningful shape. The subset contains 3,697 manually generated segmentations for 380 meshes models spread evenly amongst 19 different object categories, and has stronger consistency and higher quality.

Our quantitative evaluations are performed on 7 automatic segmentation algorithms frequently-used recently [5–10]: Randomized Cuts, Shape Diameter Function, Normalized Cuts, Core Extraction, Random Walks, Fitting Primitives and K-Means.

64.4.2 Properties Comparison of UMA, FME and AME

In this section, we perform three experiments to illustrate some properties of UMA metrics, comparing with FME and AME, and the results are shown in Figs. 64.1, 64.2 and 64.3.

The first property is the robust when deal with some segmentation in degenerative case. We provide a k-mean segmentation with random cluster center, which includes some degenerative segmentation with only 2 or 3 segments or more than 50 segments, and some random segmentation with a number similar to that of ground-truths. The **score** of three new metric are computed for this segmentation and averaged over the entire data set, Fig. 64.1 presents the experiment results. Since this segmentations is random or degenerative, so the data set is totally different from the manually ground-truth, the scores of three metrics are low, but stable for the segmentation.

The second property is the sensitivity to refinement. We compared 12 versions to the ground-truths, Fig. 64.2 illustrates that all of the three novel metrics present a slight variation (not full invariant to refinement), but still be stable.
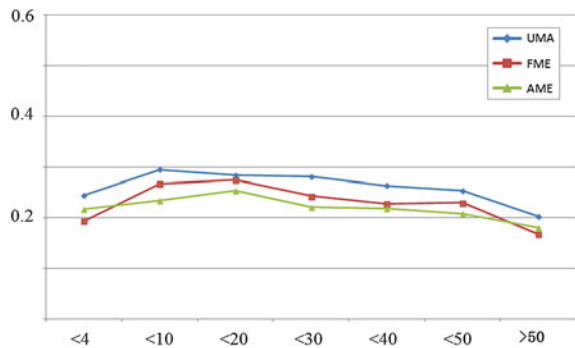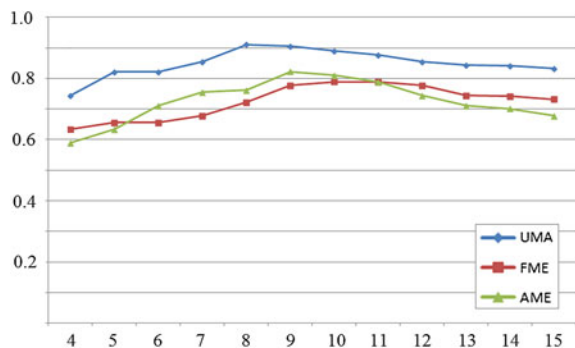
Fig. 64.1 Robust to random and degenerative cases

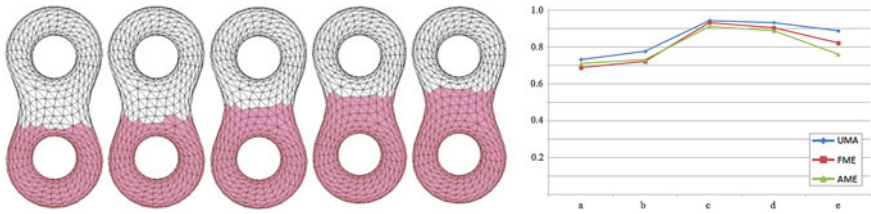**Fig. 64.2** Sensitivity to hierarchical refinement

Fig. 64.3 Tolerance to the imprecision of cut boundaries

We proposed five versions of segmentation by manually segmented the model bi-torus into two segments, and each has a slight difference to others with the boundary position. And the results in Fig. 64.3 shows that, segmentation (a) and (b) are more imprecision than the segmentation (c–e), the scores illustrate this difference indeed, and all the three novel metrics present slight variation and a good tolerance to the imprecision of cut boundaries.

### 64.4.3 Comparisons of Our Metric and Rand Index

Basing on the subset with 3,697 manually generated segmentations and according to the metrics of UMA, FME, AME and Rand Index [3], we perform other experiments also to compare the rank of the seven algorithms for each object category, the results in Table 64.1 shows that, according to the rank of UMA, the algorithm Randomized Cuts, Shape Diameter function, Normalized Cuts are on the top three, basically identical with that of Rand Index metrics, the rank of other four algorithms are similar too, this indicates that the performance of UMA metric and Rand Index metric are almost same, but FME and AME show a slight difference to them. And there is little difference between the ranks according to the metric of UMA and Rand Index too.

From the chart of UMA, FME, AME and Rand Index in Fig. 64.4, we see that each of the four bars of all four evaluation metrics is remarkably consistent with others. Although the UMA metric is boundary-based, and other three metrics focus on region dissimilarities (FME, AME and Rand Index), all variants of all four

Table 64.1 Comparison of UMA, FME, AME and Rand Index

Metrics	Rand Cuts	Shape Diam	Norm Cuts	Core Extra	Rand Walks	Fit Prim	K Means
UMA	2	3	1	4	5	7	6
FME	1	3	2	5	4	6	7
AME	1	2	3	6	4	5	7
Rand Index	1	3	2	5	6	4	7

Fig. 64.4 Comparison of FME, AME, UMA and Random Index metrics
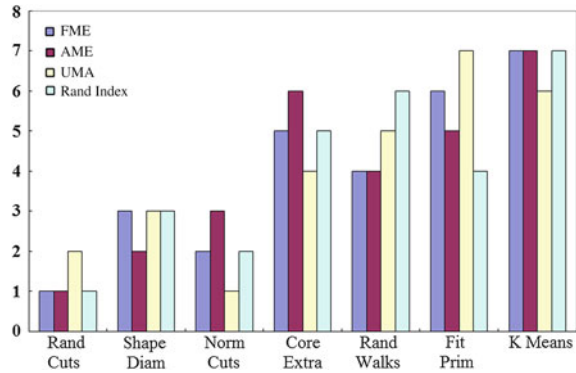
metrics suggest almost the same relative performance of the seven algorithms. The ranks of the four kinds of metrics are basically same. Because our experiments are basing a selection of the Chen's manual interactive segmentation data set, it has higher consistency, and leads to the rank of algorithms are significantly different with that in Chen et al. [3].

64.5 Conclusion

This paper presents a novel metric UMA to evaluate the algorithm of 3D mesh segmentation quantitatively, and experiments show that UMA is robust to degenerative segmentation and hierarchical refinement, stable to the imprecision of cut boundaries. Because we have excluded some manual segmentation with obvious error and serious inconsistencies, and use a data set with larger size, the UMA metric gives a more precise quantitative evaluation than others.

References

1. Shamir A (2008) A survey on mesh segmentation techniques. Comput Graph Forum 27(6):1539–1556
2. Benhabiles H, Vandeborre J, Lavoue G et al (2009) A framework for the objective evaluation of segmentation algorithms using a ground-truth of human segmented 3d-models. In: Proceedings of IEEE international conference on shape modeling and applications, Beijing
3. Chen X, Golovinskiy A, Funkhouser T (2009) A benchmark for 3D mesh segmentation. ACM Trans Graph 28(3):73:1–73:12

4. Benhabiles H, Vandeborre J, Lavoue G et al (2010) A comparative study of existing metrics for 3D-mesh segmentation evaluation. Visual Comput Int J Comput Graph 26(12):1451–1466
5. Shlafman S, Talm A, Katz S (2002) Metamorphosis of polyhedral surfaces using decomposition. Comput Graph Forum 21(3):219–228
6. Lai YK, Hu SM, Martin RR, Rosin PL (2009) Rapid and effective segmentation of 3D models using random walks. Comput Aided Geom D 26(6):665–679
7. Attene M, Falcidieno B, Spagnuolo M (2006) Hierarchical mesh segmentation based on fitting primitives. Visual Comput 3(22):181–193
8. Golovinskiy A, Funkhouser T (2008) Randomized cuts for 3D mesh analysis. ACM Trans Graphics (TOG) 5(27). doi:10.1145/1409060.1409098
9. Katz S, Leifman G, Tal A (2005) Mesh segmentation using feature point and core extraction. Visual Comput 21(8–10):865–875
10. Shapira L, Shamir A, Cohen-Or D (2008) Consistent mesh partitioning and skeletonisation using the shape diameter function. Visual Comput Int J Comput Graph 4(24):249–259
11. Sun XP, Zhao XN, Li CF et al (2012) Misclassified evaluation metrics for 3D mesh segmentation. J Chin Comput Syst 33(8):1811–1815