3D MODEL RETRIEVAL BASED ON VISION FEATURE FUSION

Mandun Zhang[1,2,3,4]，Yingshi Ma[1,2,3,4], Xiaofang Wang[1,2,3], ,Wei Wei[1,2,3,4] ,Zhidong Xiao[5]

[1] School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401，China

[2] Tianjin International Joint Center for virtual reality and visual computing, Tianjin 300401，China

[3] Hebei Province Key Laboratory of Big Data Calculation, Tianjin 300401, China.

[4] Weier Kebao(Tianjin) Science & Technology Co. Ltd.，Tianjin 300401，China

[5] National Center for Computer Animation (NCCA)，Bournemouth University，BH12 5BB，UK

zhangmandun@scse.hebut.edu.cn, yingshi.ma@outlook.com

Abstract: By applying multi-view features of a 3D model as the whole feature descriptors to match the 3D model feature, this paper presents a modified 3D model retrieval algorithm which is based on the fusion of contour features and texture features of the model. After two-dimensional depth images of the 3D model are obtained under the different views of a spherical bounding box, the contour feature and the texture features of the model images are fused for realizing the constitute of the 3D model. The experiment results shown that the proposed method gains great improvement in retrieval speed and effective rate in comparison with other view-based 3D model retrieval methods.

Keywords ： 3D model; Feature descriptor; View projection; Contour feature; Texture feature

1. Introduction

At present, according to the difference of basic unit that constitute the 3D models, the 3D model retrieval methods are mainly classified into three categories: polygon model retrieval methods[1][8], point cloud 3D model retrieval methods[9] and voxel 3D model retrieval methods[10][13]. From the internet application point of view, the usage of polygon model is greater than the other two models mentioned earlier. More and more research in the subject of 3D model retrieval focus on polygon based method, and so do this research.

For 3D polygon models, the retrieval methods are classified into four categories specifically: geometry-based retrieval methods, statistical analysis-based retrieval methods, topology-based retrieval methods and view-based retrieval methods. In these fields, many researches had been investigated and developed in the past. The multi-view based 3D model retrieval methods show the promising, and is superior to other retrieval methods. Therefore, achieving view descriptor becomes an important part in the process of 3D model retrieval. Various descriptors such as Elevation descriptor (ED) [14], Light-field descriptor (LD) [1] and Spin Image Signatures Descriptor (SISD) [15] had been proposed. In the process of LD-based retrieval, in order to improve the robustness to rotation, translation and deformation, many different view models are joined together to collect light field information. Because of the increases of view models, the number of corresponding images increase sharply that lead to decline the retrieval speed. In consideration of the fact, in order to enhance the retrieval speed, in this paper, the number of observation points are reduced and the contour features are extracted in each view. Also, the information of image color and texture might loss during the process of extracting contour of the 3D model in the corresponding view. To compensate the loss information, the whole feature points of depth image is extracted in the corresponding view at the same time, then the two features in the same view are fused as the feature descriptor of the corresponding view.

2. Model preprocessing

For 3D models, the projection outcome is affected by

translation, rotation, scale and position. To extract the contour and texture features from 3d model projection images, the 3D model pre-processing is necessary to ensure the model invariant. Four steps are aligned with pre-processing to ensure the invariance. The first step: a model is placed at the origin of coordinate systems to ensure translation invariance. The second step: PCA is used to align models in a standard coordinate plane to ensure rotation invariance [13]. The third step: flipping transformation is needed to ensure position invariance. The fourth step: the model is normalized into a standard unit to ensure scale invariance. Therefore, the standardization process for constructing a whole coordinate model can be expressed as follows:

$$\tau(I) = S^{-1} \cdot F \cdot R \cdot (I - c) \qquad (1)$$

Where S is the scaling coefficient, F is a reversal matrix in the form of diagonal matrix, R is the rotation matrix obtained by PAC transformation, I is the original model coordinate, c is the coordinate origin. Figure 1 is the transform result of fish model by PCA.
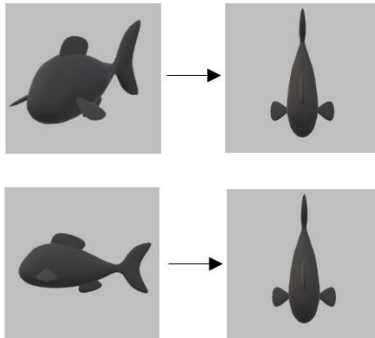

Figure 1 Transform result after PCA

3. Mixed model descriptor

The steps to build a mixed model descriptor are as follows: sphere is set as the bounding box and the center of the sphere is set as the origin of the coordinate, the 3D model is put in the center of the sphere, every observation point is set on the equatorial plane of the spherical surface every 45 degrees, in this way there are total 8 observation points on the equatorial plane, meanwhile the two poles of the spherical surface are selected as top and bottom observation points. the intersection of the spherical surface and the positive direction of the X axis is set as the initial position, these eight points are encoded respectively

as $\omega_1 \sim \omega_8$ by the counter-clockwise direction of the right-hand rule, ω_0, ω_9 represent respectively the top and bottom poles as shown in Figure 2.
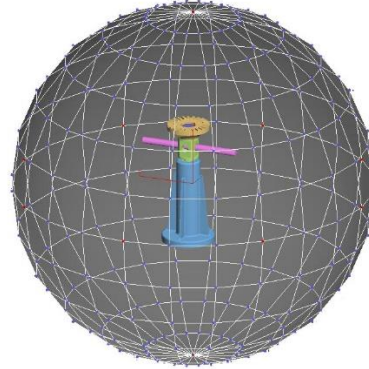

Figure 2 Sphere bounding box

Firstly, the two-dimensional depth image of the model is obtained at the observation position ω_i, thus the contour image and contour feature L_i are able to be achieved, and at the meantime the texture feature T_i is able to be obtained by extracting the feature of 2D depth image. Therefore, the whole feature of 3D model in this view can expressed by ω_i, L_i, T_i:

$$F = \sum_{i=0}^{9} F_i \qquad (2)$$

$$F_i = (\omega_i, L_i, T_i)$$

where F_i is the mixed model descriptor of the contour feature L_i and the texture feature T_i in the view ω_i, F is the feature vector group that generated under different views.

4. Feature extraction and feature description

4.1 Feature extraction

For 3D models, model placement and different topological changes can cause different self-occlusion problems, so the contour of the model from the projection is especially important in the view-based retrieval method. In order to obtain the accurate contour image of the model, an improved Canny operator is applied to extract the contour feature L_i of the model and the surf operator is used to extract the texture feature T_i of the model, then the two features are fused to obtain the whole feature of the model F_i.

The traditional Canny operator is more accurate in locating edges and is sensitive to noise. It is easy to detect the false edge and some distortion of the edge details. The problem is more prone to produce false edge while dealing with self-occlusion part of 3D model. Therefore, this paper presents a modified method for determining the magnitude of pixel gradient in the 8 pixel neighborhood by calculating the first order partial derivative of the X direction, the Y direction, the direction of the 135° and the 45° direction. A finite difference method is applied to calculate the gradient in each direction.

Partial derivative of X direction:
$$P_x[i,j] = I[i+1,j] - I[i-1,j]) \qquad (3)$$

Partial derivative of Y direction:
$$P_y[i,j] = I[i,j+1] - I[i,j-1] \qquad (4)$$

Partial derivative of 45° direction:
$$P_{45°}[i,j] = I[i+1,j+1] \\ - I[i-1,j-1] \qquad (5)$$

Partial derivative of 135° direction:
$$P_{135°}[i,j] = I[i-1,j+1] \qquad (6) \\ - I[i+1,j-1]$$

Where X and Y are row and column direction, respectively.

The contour of the model is obtained through the calculation of the gradient, and then the feature points of the edges are detected by using Harris corner detector. Therefore, the main direction is able to be defined and feature vector descriptors L_i for each feature point are able to be generated to improve the efficiency of edge feature matching. At the same time, the Surf algorithm is used for texture feature extraction under the same view point, and the Hessian matrix is:

$$H(x, \; \sigma) = \begin{bmatrix} L_{xx}(x, \; \sigma) & L_{xy}(x, \; \sigma) \\ L_{xy}(x, \; \sigma) & L_{yy}(x, \; \sigma) \end{bmatrix} \qquad (7)$$

The local texture feature T_i is extracted by Surf algorithm, and is spliced with the global contour feature L_i to generate the Local and global feature vector F_i, the final model feature vector group F with multiple perspectives is constructed in the end. In our research experiment, 3 model libraries had been applied for testing our algorithm: Mc-Gill (McGill University benchmark database), PSB (Princeton shape benchmark) and ESB (Purdue Engineering). The model in McGill can extract about 200 global contour features and 350 local features. Each model in the PSB can extract about 200 global features and 250 local features. Each model in ESB can extract about 100 global features and 150 local features.

4.2 feature classification

First, K-means++ algorithm is utilized for model clustering for the 3 different model libraries. After 20 iterations on the cluster data, we are able to obtain m classifications. The number of global contour features and local Surf features in each class are able to be counted. A m dimensional histogram is able to be generated and is able to be transformed into a m dimensional vector for representing the feature vector of the 3D model. Furthermore, both global and local features are able to be clustered into m classifications. By counting the number of the feature vectors of each classification, a histogram is able to be generated which is able to be used for forming feature vectors with global and local features.

5.   Similarity matching

Because both contour features and texture features are included in the feature vectors group F, meanwhile F_i represents the feature vector under the view point i. When calculating the similarity between the models, Euclidean Distance is adopted to calculate the distance between the feature vectors in the experiments. For the above feature vector group F, D represents the distance between the feature vector groups of two 3D models:

$$D = \sqrt{|F^m - F^n|^2} \\ = \sqrt{\sum_{i=0}^{9}(F_i^m - F_i^n)^2} \qquad (8)$$

where m and n represents different number of model.

Then the distance between any two 3D models is calculated, and the similarity measurement between the models is obtained.

6.   Experimental results and analysis

The experimental results in Table I indicate that the relationship between feature types, number of features and number of clustering centers in McGill model database.

Table I The number of mixed features

Model type	Number of features (N_1)	The number of clustering center (k_i)
Ant	12779	121
Crab	15145	130
Hand	3578	35
Human	7967	69
Cuttle	2455	27
pliers	3420	34
Snake	3120	32
Glasses	4418	39
Bug	13348	127
Bear	4656	41
Plane	6050	45
Bird	3122	32
Chair	5640	43
Cup	2330	25
Dinosaur	3950	38
Fish	7602	50
Table	4394	40

By applying the same method to cluster the models in the PSB database and the ESB database, we are able to build the mixed feature descriptors. To evaluate the effectiveness of this method in 3D model retrieval, the precision – recall rate from 3 different model libraries experiment is shown in Figure 3.
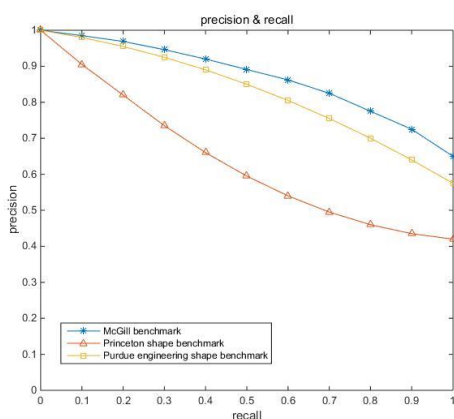

Figure 3 Precision and recall of different benchmark

In our research experiment, our presented multi-view based global and local feature fusion algorithm had been compared with, D2 (D2 shape distribution) algorithm, MODFD (multiple orientations depth Fourier descriptor) algorithm and AAD (absolute angle distance histogram)

algorithm. By calculating the Euclidean distance of two random points, D2 algorithm presents a shape function which statistically describes the geometric distribution of the model. The MODFD algorithm employs orthogonal projections of a three-dimensional model in 42 view-points. By calculating the universal Fourier descriptors of each orthogonal projection from depth map, the feature descriptors of the three-dimensional model are constructed. The AAD algorithm is a 3D model feature extraction algorithm which is based on shape. Firstly, a 3D model with surface information is converted into a directed point set model, and then, the calculation of the distance between all pairs of points is carried out. The calculated distances are to form a 2D histogram. Same model libraries had been tested with the three mentioned algorithms, the comparison results of precision and recall rate are shown from Figure 4 to Figure 6.
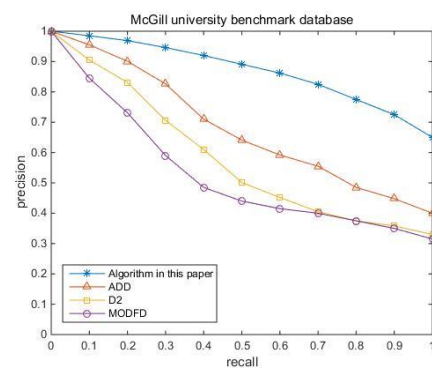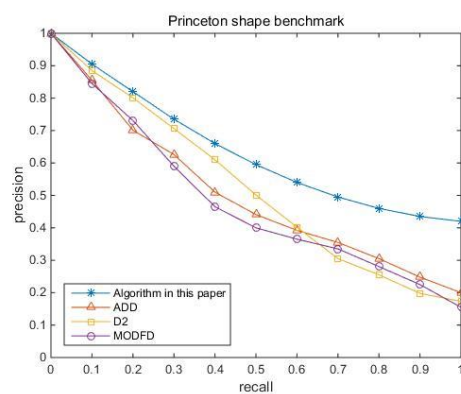

Figure 4 Precision and recall on McGill

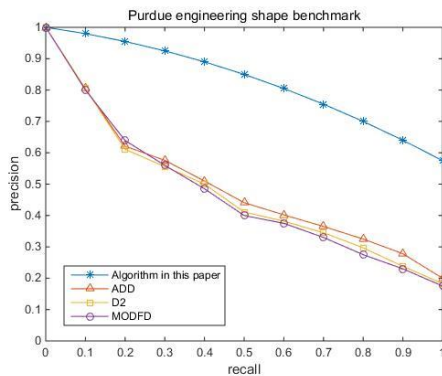**Figure 5** Precision and recall on PSB

Figure 6 Precision and recall on ESB

7. Conclusion

In this paper, we present a 3D model retrieval algorithm based on the fusion of global features and local feature from multi-perspective depth map. This algorithm transforms the complex 3D model extraction problem into a simple 2D image feature extraction problem through multiple views. Thus, the contour feature extraction algorithm in the image processing domain and the local feature extraction algorithm for image key points can be used to extract the model features. These global and local features are quantized and clustered to generate histograms and expressed as feature vectors to describe the 3D models. This not only ensures the integrity of the feature description, but also achieves low complexity. At the same time, the algorithm proposed is compared with the various existing algorithms such as D2, MODFD and AAD etc. The experimental results show that the algorithm presented in this paper has advantages in precision and recall rate. In recent years, algorithms in the deep learning field are becoming more and more popular and promising, to enhance and gain further improving in the success rate of 3D model retrieval, a deep learning method for extracting 3D model features may be considered in the future research.

Reference

[1] Chen, D. Y., Tian, X. P., Shen, Y. T., & Ouhyoung, M. On visual similarity based 3D model retrieval. Computer graphics forum, 2003, 22(3): 223-232.

[2] Eitz, M., Hildebrand, K., Boubekeur, T., & Alexa, M. Sketch-based 3D shape retrieval. SIGGRAPH Talks, 2010, 3.

[3] Dutagaci, Helin, Chun Pan Cheung, and Afzal Godil. A benchmark for best view selection of 3D objects. Proceedings of the ACM workshop on 3D object retrieval. ACM, 2010:45-50.

[4] Eitz, M., Hays, J., & Alexa, M. How do humans sketch objects?. ACM Trans. Graph., 2012, 31(4): 44-1.

[5] Wang, F., Kang, L., & Li, Y. Sketch-based 3d shape retrieval using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), 2015: 1875-1883.

[6] Li, B., Lu, Y., Li, C., Godil, A., Schreck, T., Aono, M. Burtscher, H. Fu, T. Furuya, H. Johan, J. Liu, R.Ohbuchi, A.Tatsuma, C.Zou SHREC'14 track: Extended large scale sketch-based 3D shape retrieval. In Eurographics workshop on 3D object retrieval, (Vol. 2014).

[7] XIAO Qin-han, LIU Cong, & Du Yong-jun.3D object retrieval based on lightfied hybrid descriptor. Computer Engineering and Applications,2010,46(2):188-194.

[8] Chaouch, M., & Verroust-Blondet, A. A new descriptor for 2D depth image indexing and 3D model retrieval. Proceedings of the IEEE International Conference on Image Processing（ICIP 2007），Vol 6：VI-373.

[9] LIU Wei. Feature extraction and retrieval of 3D model. Doctoral dissertation, Shanghai Jiao Tong University，2008.

[10] Xie, J., Zheng, Z., Gao, R., Wang, W., Zhu, S. C., & Wu, Y. N. Learning Descriptor Networks for 3D Shape Synthesis and Analysis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：8629-8638.

[11] Chen Ruiqing. 3D feature extraction and matching based on voxel. Doctoral dissertation, Shanghai Jiao Tong University，2015.

[12] Leng Biao. Research on feature description and relevance feedback algorithm of 3D model retrieval. Doctoral dissertation, Tsinghua University，2009.

[13] Cui Chenchang, & Shi Jiaoying. Research on feature description and relevance feedback algorithm of 3D model retrieval. Journal of Computer-Aided Design and Computer Graphics, 2004, 16(7): 882-889.

[14] Shih, J. L., Lee, C. H., & Wang, J. T. A new 3D model retrieval approach based on the elevation descriptor. Pattern Recognition, 2007, 40(1): 283-295.

[15] Bimbo A D，Pala P.Content-based retrieval of 3D

models.ACM Trans on Multimedia Computing，2006,2(1):20-43.