Improved 3D Local Feature Descriptor Based on Rotational Projection Statistics and Depth Information

Hui Zeng(✉), Rui Zhang, and Mingming Huang

School of Automation and Electrical Engineering,
University of Science and Technology Beijing, Beijing 100083, China
hzeng@163.com

Abstract. 3D local feature descriptor construction is a very challenging task in the field of 3D model analysis. In this paper, an improved Rotational Projection Statistics (IRoPS) descriptor is proposed. For each feature point, the local coordinate system is firstly built and its neighboring points are normalized. Then the normalized neighboring points are rotated and projected onto three coordinate planes. For each rotation, the distribution matrix is computed and the sub-descriptor can be obtained using the central moment, the Shannon entropy, the mean and variance of local depth values. Finally the IRoPS descriptor is constructed by concatenating all the sub-descriptors into a vector. Compared with the Rotational Projection Statistics (RoPS) descriptor, the IRoPS descriptor includes the local depth information and it has better discriminative power. Extensive experiments are performed to verify the superior performance of the proposed descriptor.

Keywords: 3D model matching · Rotational projection statistics · Local feature descriptor · Depth information

1 Introduction

As the increasing development of the computer vision theory, 3D model has gradually become the fourth type of multimedia data following the sound, image and video. Compared with image, 3D model contains much more information, which could be more conducive to the analysis and understanding of the scene. Therefore, 3D model has been applied in more and more fields, such as virtual reality, 3D games, building design, animation effect and medical diagnosis. Especially in recent years, with the emergence of 3D printer, the application of 3D model has begun to spread to households, which making the home users can print 3D model [1]. As a result, how to quickly search for the required model from so many 3D models has become an urgent research topic.

For most 3D model retrieval methods, 3D model matching is one of the basic steps [2]. Therefore, this paper mainly investigates the construction of the 3D model local descriptor and its application in 3D model matching. In general, model matching

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 139–147, 2015.
DOI: 10.1007/978-3-662-48570-5_14

includes the following process. Firstly, the feature descriptor is extracted from the 3D model. Then the similarity between the original model and the corresponding model is calculated to accomplish the 3D model matching. Up to now, according to the different extracted methods of feature descriptor, there are two kinds of methods to solve this problem: global matching methods and local matching methods. Global matching methods represent global shape of the entire 3D model as a feature descriptor, such as geometric distribution based methods [3], skeleton graph matching based methods [4] and projection map based methods [5], etc. However, when 3D models have similar local structure, pose variations and partial occlusions, global matching is difficult to achieve better performance. Over the past two decades, the most popular trend for 3D model matching is to use local features, due to their better robustness and descriptiveness. The local features contain abundant shape information, and they are not easily influenced by external environment, such as pose and lighting variations, shape deformation and occlusions. Local matching methods mainly have two kinds of methods. The first ones are feature points description based local matching methods. The keypoints are detected and their local descriptors are constructed for matching. This kind of methods is relatively robust to noise, but the global topology information is not been fully used in the process of matching. The second ones are topology description based local matching methods. This kind of methods obtains the corresponding local geometry features for sub-graph matching. But the normal vector and curvature calculation would be existed in descriptor construction, resulting that this kind of methods would be less robust to noise.

From the above analysis we can see that the construction of 3D local descriptor is one of key research topics of 3D local matching methods. So, in this paper, our research emphasis is 3D local descriptor construction. Up to now, much work about 3D local descriptor construction has been published. Chua et al. proposed point signature based feature descriptor [6]. They obtained a contour C by intersecting the 3D model surface with a sphere of radius r centered at the keypoint p. Then, they fitted a plane to these contour points. The distances between contour points and fitting plane would be used for descriptor construction. Johnson et al. proposed spin image based feature descriptor [7]. They used the normal vector of a keypoint p as the local reference axis to build a cylindrical coordinate system. A spin image based feature descriptor is generated by projecting a local surface onto the 2D cylindrical plane. Mian et al. proposed tensor based feature descriptor [8]. Firstly, local points were preprocessed by triangular mesh method. They then constructed a local 3D grid and summed the surface areas in each bin of the grid, to generate a "3D tensor" descriptor. Recently, Yulan Guo et al. proposed a novel local feature extraction algorithm, named Rotational Projection Statistics (RoPS) [9], and its performance is better than five state-of-art descriptors, including spin image [7], normal histogram (NormHist) [10], Local Surface Patch (LSP) [11], THRIFT [12] and signature of histograms of orientations (SHOT) [13]. The RoPS descriptor is generated by rotationally projecting the neighboring points onto three local coordinate planes and calculating several statistics (central moment and Shannon entropy) of the projected points, showing both high discriminative power and strong robustness to noise. However, this method only focused on the distribution statistics of the feature points, without considering the depth information of the projected points.

In this paper, in order to make full use of 3D models' local depth information, we propose an improved local feature descriptor, named improved Rotational Projection Statistics (IRoPS) descriptor. The IRoPS descriptor use the frame of RoPS descriptor, and the central moment and the Shannon entropy are also used for the sub-descriptor construction. The improvement of the IRoPS descriptor is that the mean and variance of local depth values are added in the process of sub-descriptor construction. Compared with the RoPS descriptor, the IRoPS descriptor can depict the 3D local structure more comprehensively. For two different local neighborhoods with different local depth information, they may have same RoPS descriptor and have different IRoPS descriptors. Comparative experiments have been performed and the results demonstrate the effectiveness and efficiency of our proposed IRoPS descriptor compared with RoPS descriptor.

The rest of this paper is organized as follows. Section 2 provides the description of the improved RoPS descriptor. Section 3 presents the results and analysis of 3D model matching experiments. Section 4 concludes this paper.

2 Improved Local Surface Descriptor construction

The construction of the improved RoPS descriptor includes the two following steps. Firstly, the 3D local coordinate system of each 3D feature point is built, which can provide invariance to 3D translation and rotation. Then the local feature descriptor is constructed, using the central moment, the Shannon entropy, the mean and variance of depth values of the projected points.

2.1 3D Local Coordinate System Definition

In this paper, we use the local reference frame proposed in [9] to build unique 3D local coordinate system for every 3D feature points, which can provide invariance to 3D rotation and transformation for the following local descriptor construction. For every feature point p, given a support radius r, the local neighborhood can be determined by cropping a sphere of radius r centered at p. For the i_{th} triangle with three vertices p_{i1}, p_{i2} and p_{i3}, the scatter matrix C_i can be defined as:

$$C_i = \tfrac{1}{12}\sum_{j=1}^{3}\sum_{k=1}^{3}(p_{ij}-p)(p_{ik}-p)^T + \tfrac{1}{12}\sum_{j=1}^{3}(p_{ij}-p)(p_{ij}-p)^T \tag{1}$$

Assume N is the number of triangles, the overall scatter matrix C is defined as:

$$C = \sum_{i=1}^{N} w_{i1} w_{i2} C_i \tag{2}$$

where $w_{i1} = \dfrac{|(p_{i2}-p_{i1})\times(p_{i3}-p_{i1})|}{\sum_{i=1}^{N}|(p_{i2}-p_{i1})\times(p_{i3}-p_{i1})|}$ $w_{i2} = (r - |\, p - \dfrac{p_{i1}+p_{i2}+p_{i3}}{3}|)^2$.

Then the descending eigenvectors $\{v_1, v_2, v_3\}$ of C are calculated, which offer a basis for local coordinate system definition. In order to eliminate the sign ambiguity, the unambiguous vector v_1 is defined as:

$$v_1 = v_1 \cdot sign(h) \tag{3}$$

where $h = \sum_{i=1}^{N} w_{i1} w_{i2} (\frac{1}{6} \sum_{j=1}^{3} (p_{ij} - p) v_1)$. Similarly, we can get the unambiguous vector v_2 and v_3. Consequently, 3D local coordinate system for given point p is finished, that p is the origin, v_1, v_2 and v_3 are the x, y and z axes respectively.

2.2 Improved RoPS Descriptor Construction

The RoPS (Rotational Projection Statistics) descriptor is a novel 3D local descriptor proposed by Yulan Guo et al. And it can be generated by rotationally projecting the neighboring points around a feature point onto three coordinate planes and calculating the statistics of the distribution of the projected points. Although extensive experiments have testified its superior performance, it does not make full use of the local structure information of the 3D point's neighborhood. Only the number of projecting points in each bin is used for the RoPS descriptor construction, without considering the depth information of the projection points. For two different local neighborhoods with same RoPS descriptor, they may have different local depth variance. The local depth information of the 3D local neighborhood is important for depicting 3D local structure. So we propose an improved RoPS descriptor to make the descriptor contains the depth information.

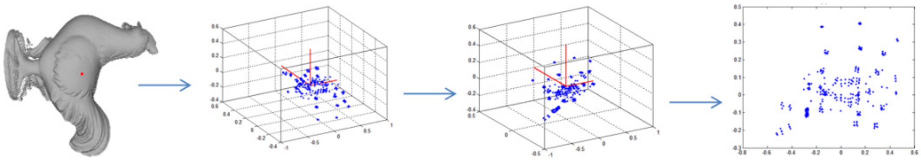
The construction process of a RoPS descriptor is shown in Fig. 1. Given a feature point p and its support radius r, local neighborhood of the feature point can be extracted and its local coordinate system can be determined using the method given in section 2.1. After rotation and translation normalization, the local neighborhood is rotated around every axis by an angle $2\pi/T$ each time, where T denotes the number of rotations. Then the local neighboring points are projected on xy plane xz plane and yz plane respectively. For each projection, the bounding rectangle of the projected points can be divided into $L \times L$ bins, and the number of points falling into each bin is counted to yield an $L \times L$ distribution matrix D. The distribution matrix D is then normalized to let the sum of its elements be 1, and it can make the descriptor have invariance to variations in mesh resolution. Then the central moment and the Shannon entropy are computed from the matrix D. The central moment μ_{mn} of the matrix D can be calculated using the following equation:

$$\mu_{mn} = \sum_{i=1}^{L} \sum_{i=1}^{L} (i - \bar{i})^m (j - \bar{j})^n D(i, j) \tag{4}$$

where, $\bar{i} = \sum_{i=1}^{L} \sum_{j=1}^{L} iD(i, j)$, $\bar{j} = \sum_{i=1}^{L} \sum_{j=1}^{L} jD(i, j)$. The Shannon entropy e of the matrix D can be calculated as:

$$e = -\sum_{i=1}^{L}\sum_{j=1}^{L} D(i,j)\log(D(i,j)) \tag{5}$$

For each rotation and projection, the sub-descriptor can be obtained using $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e\}$. Finally the RoPS descriptor can be obtained by concatenating all the sub-decriptors into a vector.



(a) 3D model (b) local neighborhood (c) rotated points (d) projected plane

Fig. 1. An illustrative example of the RoPS method.

The RoPS method represents local points from a set of views, converting the matching between 3D models to that between the 2D projected planes. Under this frame, many 2D image matching methods can be used for 3D model matching. However, the difference between 3D model and 2D image is that 3D model has abundant depth information. The RoPS descriptor only uses the distributed information of the projected points from different views. The points with similar position in projected plane would have different depths in 3D local coordinate system. Unavoidably, the RoPS descriptor loses part of the 3D spatial structural information. In the history of machine vision, psychologist pointed that the human visual system uses a lot of depth information based on visual sense to understand and identify objects [14]. This give an important enlightenment: computer vision researchers can directly investigate the visual system based on depth information. As we all know, snapshot descriptor is obtained by using depth information. Malassiotis and Strintzis [19] first constructed an 3D local coordinate system by performing an eigenvalue decomposition on the covariance matrix of the neighboring points of a keypoint p. They then placed a virtual pin-hole camera at a distance d on the z axis and looking toward p. The x and y axes of the camera coordinate frame were also aligned with the x and y axis of the 3D local coordinate system at p. They projected the local surface points onto the image plane of the virtual camera and recorded the distance of these points from the image plane as a ”snapshot” descriptor. The snapshot descriptor is robust to self-occlusion and very efficient to compute. Snapshot achieved better pairwise range image alignment results compared to spin image. Mian et al. [20] also defined an 3D local coordinate system for a local surface and then fitted the local surface with a uniform lattice. They used depth values of the local surface to form a feature descriptor, which was further compressed using a PCA technique.

Inspired by this conclusion and some previous work, we improve the RoPS descriptor using the statistic of depth information. For each distribution matrix D, its mean and variance of depth information can be defined as:

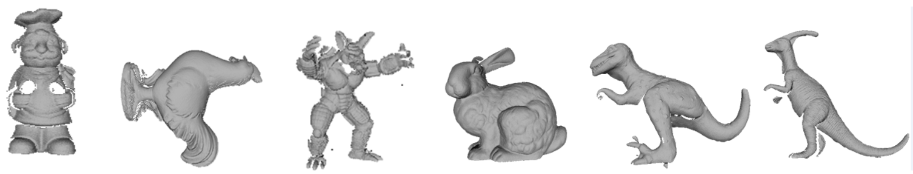
$$d_m = \frac{\sum_{i=1}^{num} |d_i|}{num} \tag{6}$$

$$d_v = \frac{\sum_{i=1}^{num} (|d_i| - d_m)^2}{num} \tag{7}$$

where d_i is the i_{th} point's depth value, *num* is the points' number. The mean value $\mathrm{d_m}$ can measure the distance between the local point set and its center point, and the variance value $\mathrm{d_v}$ can measure how far the local points are spread out. Therefore, they are useful for describing the 3D local structure. Finally, the sub-descriptor of each projection can be obtained using $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e, d_m, d_v\}$ and the improved RoPS descriptor can be generated by concatenating all the sub-descriptors into a vector.

3 Experimental Results

In this section, comparative experiments are performed to verify the superiority of the proposed method. As shown in Fig. 2, the experimental data are six models ("Armadillo", "Bunny", "Chicken", "T-rex", "Parasaurolophus" and "Chef"), which were taken from the Stanford 3D Scanning Repository [15] and Mian's Dataset [16,17]. The corresponding model was synthetically generated by randomly rotating in order to create clutter and pose variances. Then Gaussian noise with a standard deviation of 0.1 mesh resolution was added to the model.

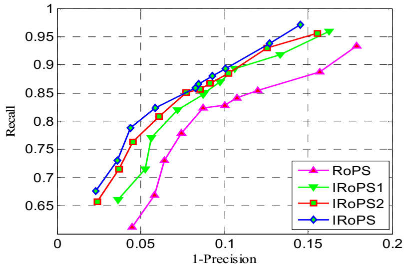


(a) Chef (b) Chicken (c) Armadillo (d) Bunny (e) Parasaurolophus (f) T-rex

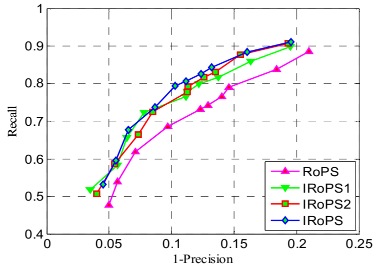
Fig. 2. The experimental data.

In the experiments, we use the parameters according to Guo's suggestion in [8], which means the parameters are $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e\}$, L=5, r=15mr (mesh resolution) and T=3. At first, 1000 feature points are randomly selected from the 3D model. Then the local descriptors are constructed for each feature point. Finally the feature points of two 3D models are matched by computing the distances of descriptors. To evaluate the effectiveness of the depth information, we compare four kinds of statistics to consruct the local descriptors.

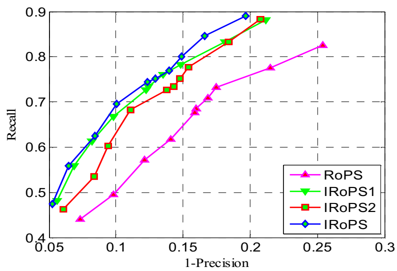
The experimental results are evaluated using the Recall-Precision criteria based on the number of the correct matches and the number of the false matches. Fig. 3 is the matching results of the testing 3D models. Here "RoPS" denotes the descriptor proposed by Yulan Guo et al. and it is constructed by $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e\}$,"IRoPS1" denotes the descriptor constructed by $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e, d_m\}$, "IRoPS2" denotes the descriptor constructed by $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e, d_v\}$, and "IRoPS" denotes the descriptor constructed by $\{\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, e, d_m, d_v\}$. The goal in PR space is to be in the upper-left-hand corner, which means the better performance in 3D model matching [18]. From Fig. 3 we can see that the performance of the IRoPS1 descriptor is better than the performance of the IRoPS2 descriptor, and the IRoPS descriptor performs best. That is to say, the mean of the depth values has more discriminative information than the variance of the depth values for the local descriptor construction, and the IRoPS descriptor can character the local structure of the 3D model effectively.



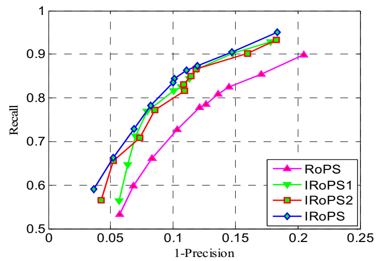
(a) Chef



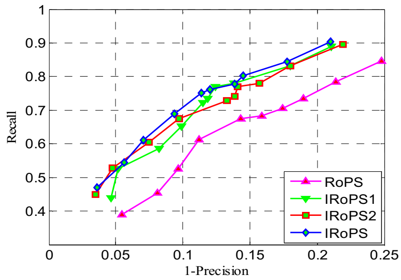
(b) Trex



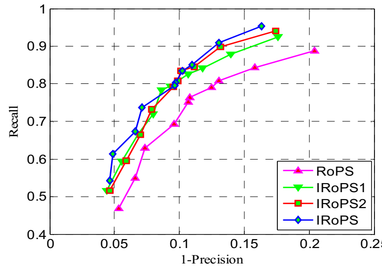
(c) Armadillo



(d) Bunny



(e) Parasaurolophus



(f) Chicken

Fig. 3. The experimental results.

4 Conclusion

In this paper, a novel 3D local feature descriptor called IRoPS is proposed. It is the improvement edition of the RoPS descriptor. Similar to the RoPS descriptor, the IRoPS descriptor is invariant to rotation and translation by normalization using the local coordinate system. Compared to RoPS descriptor, the IRoPS descriptor not only contains the distribution information of the projected points but also contains the local depth information. It has better discriminative power, while maintaining the robustness of the RoPS descriptor. To verify the performance of the proposed descriptor, extensive 3D model matching experiments have been performed. The experimental results show that the IRoPS descriptor better performance than the RoPS descriptor. In our future work, we will search more discriminative information of the 3D local surface to improve the performance of the descriptor.

Acknowledgments. This article is supported by the National Natural Science Foundation of China (Grant No. 61375010, No.61005009), Beijing Higher Education Young Elite Teacher Project (Grant No.YETP0375) and the Fundamental Research Funds for the Central Universities under the Grant No. FRF-TP-14-120A2. The authors would like to acknowledge Yulan Guo for providing the code of RoPS descriptor.

References

1. Waran, V., Narayanan, V.: Utility of multimaterial 3D printers in creating models with pathological entities to enhance the training experience of neurosurgeons: Technical note. Journal of Neurosurgery **120**(2) (2014)
2. Wang, M., Gao, Y., Lu, K., Rui, Y.: View-based discriminative probabilistic modeling for 3d object retrieval and recognition. IEEE Transactions on Image Processing **22** (2013)
3. Ohbuchi, R., Otagni, T., Ibato, M., et al.: Shape similarity search of three-dimensional models using parameterized statistics. In: Proceedings of the Pacific Graphics, pp. 265–274. Beijing, China (2002)
4. Sundar, H., Silver, D., Gagvani, N., et al.: Skeleton based shape matching and retrieval. In: Shape Modeling International, pp. 130–139. IEEE (2003)
5. Chen, D.-Y., Tian, X.-P., Shen, Y.-T., et al.: On Visual Similarity Based 3D Model Retrieval. Eurographics V22(3) (2003)
6. Chua, C., Jarvis, R.: Point signatures: a new representation for 3Dobject recognition. International Journal of Computer Vision **25**(1), 63–85 (1997)
7. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(5), 433–449 (1999)
8. Mian, A., Bennamoun, M., Owens, R.: Three-dimensional model-based object recognition and segmentation in cluttered scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(10), 1584–1601 (2006)
9. Guo, Y., Sohel, F., Bennamoun, M., Min, L., Wan, J.: Rotational Projection Statistics for 3D Local surface description and object Recongnition. Int. J. Comput. Vis. **105**, 63–86 (2013)

10. Hetzel, G., Leibe, B., Levi, P., Schiele, B.: 3D object recognition from range images using local feature histograms. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, issue II, p. 394 (2001)
11. Chen, H., Bhanu, B.: 3D free-form object recognition in range images using local surface patches. Pattern Recognition Letters **28**(10), 1252–1262 (2007)
12. Flint, A., Dick, A., Van den Hengel, A.: Local 3D structure recognition in range images. IET Computer Vision **2**(4), 208–217 (2008)
13. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: European Conference on Computer Vision, pp. 356–369. Crete, Greece (2010)
14. Okoshi, M.T.: Depth Cues in the Human Visual System, Three-Dimensional Imaging Techniques. Academic Press, New York (1976)
15. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: 23rd Annual Conference on Computer Graphics and Interactive, Techniques, pp. 303–312. New Orleans, LA (1996)
16. Mian, A.S., Bennamoun, M., Owens, R.: A Novel Representation and Feature Matching Algorithm for Automatic Pairwise Registration of Range Images. International Journal of Computer Vision **66**(1), 19–40 (2006)
17. Mian, A., Bennamoun, M., Owens, R.: 3D Model-based Object Recognition and Segmentation in Cluttered Scenes. IEEE Transactions in Pattern Analysis and Machine Intelligence **28**(10), 1584–1601 (2006)
18. Davis, J., Goadrich, M.: The relationship between precision-recall and ROC curves. In: Proceedings of 23rd International Conference on Machine Learning, Pittsburgh, PA (2006)
19. Malassiotis, S., Strintzis, M.: Snapshots: A novel local surface descriptor and matching algorithm for robust 3D surface alignment. IEEE Trans. Pattern Anal. Mach. Intell. **29**(7), 1285–1290 (2007)
20. Mian, A., Bennamoun, M., Owens, R.: On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. Int. J. Comput. Vis. **89**(2), 348–361 (2010)