3D Part Identification Based on Local Shape Descriptors

Xiaolan Li[1,2] Afzal Godil[1] Asim Wagan[1]

[1] National Institute of Standard and Technology, Gaithersburg, MD 20899, U.S.A.

[2] Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China

Contact: lixlan@nist.gov

ABSTRACT

This paper explores 3D object recognition based on local shape descriptor. 3D object recognition is becoming an increasingly important task in modern applications such as computer vision, CAD/CAM, multimedia, molecular biology, robotics, and so on. Compared with general objects, CAD models contain more complicated structures and subtle local features. It is especially challenging to recognize the CAD model from the point clouds which only contain partial data of the model.

We adopt the Bag of Words framework to do the partial-to-global 3D CAD retrieval. In this paper the visual words dictionary is constructed based on the spin image local feature descriptor. The method is tested on the Purdue Engineering Benchmark. Furthermore, several experiments are performed to show how the size of query data and the dissimilarity measurement affect the retrieval results.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Shape, Representations, data structures, and transforms.

General Terms

Algorithms, Performance, Reliability.

Keywords

CAD model retrieval, bag of words, spin image.

1. INTRODUCTION

Large number of 3D models are created everyday and stored in databases. In order for these 3D databases to be useful, we should be able to search on them. Therefore, identification, retrieval and classification of 3D objects are becoming an increasingly important task in modern applications such as computer vision, CAD/CAM, multimedia, molecular biology, robotics, and so on.

With recent developments in 3D range scanners it is possible to capture 3D shapes in real time. However, because of the limitation of the point of view, the occlusion in the scene, and the real time requirement, only parts of the object can be captured during scanning. This proposes a challenging research problem: given an incomplete point cloud of an object, how to retrieve the corresponding complete model from a database. Solving this problem will also benefit several other applications, such as data registration [Mitra06], model fixing [Founkhouser04], and so on.

Nevertheless, most of the 3D shape retrieval methods are based on global shape descriptors, which require the complete geometry of a 3D object, such as Light Field descriptors [Chen03], spherical harmonics descriptor [Kazhdan03], D2 shape distribution [Osada02]. That these methods are not suitable for solving the problem provides an impetus to create methods for partial-to-global 3D shape identification and matching.

Besides the benefits of partial-to-global retrieval, local descriptors can capture more local details than can the global ones. Compared with general objects, CAD models have more complicated structure with holes and other local features. Using global information, these subtle details can be neglected. From this aspect, local descriptors are better.

In this paper, we present a complete framework for performing 3D partial shape identification on 3D CAD parts. Several experiments are performed to show how the size of query data and the dissimilarity measurement affect the retrieval results.

The organization of the paper is as follows. Several related works are summarized in Section 2. Section 3 outlines the whole framework of our method, and introduces two crucial terms: bag-of-words and spin image. Then, the procedures of feature extraction and similarity computation are described in Section 4. In Section 5, we provide the 3D shape retrieval results on the Purdue Engineering Benchmark.

2. RELATED WORK

In order to perform 3D partial-to-global shape retrieval, the following methods have been proposed. [Podolak06] exploits the symmetry of the shape. [Mitra06] [Frome04] develop local shape signatures. Because of its simplicity and generality, the bag-of-words method, which is insensitive to deformation, articulation and partial missing data, has attracted lots of interest in 2D [Li05] and 3D [Shan06] [Liu06] [Ohbuchi08] fields. In [Li05], the method is applied to images by using a visual analogue of a word, formed by vector quantizing two regional descriptors: normalized 11*11 pixel gray values and SIFT descriptors. In [Shan06] and [Liu06], visual feature dictionary is constituted by clustering spin images in small regions. In order to procure partial-to-whole retrieval, Kullback-Leibler divergence is proposed as similarity measurement in [Liu06], while a probabilistic framework is introduced in [Shan06]. For the sake of collecting visual words, Ohbuchi et. al. [Ohbuchi08] apply SIFT algorithm to depth buffer images of the model captured from

PerMIS'08, August 19–21, 2008, Gaithersburg, MD, USA.

uniformly sampled locations on a view sphere. After vector quantization, Kullbak-Leibler divergence measures the similarities of the models. But these methods focus on the retrieval of general objects.

Compared with general objects, CAD models have a more complicated topology with holes and other local features. In [Ip07], partial CAD retrieval is achieved based on segmentation, which directly affects the retrieval results. This paper aims to develop a new method for 3D CAD parts identification in similar circumstance as in [Ip07]. That is, given an unknown partial 3D point cloud of a part, we are trying to identify the part based on the known CAD model in a database. Moreover, our framework is closely related to that of [Liu06] and [Shan06], which does not require segmentation at all.

3. OUR FRAMEWORK

We first describe the whole framework of our method, and then introduce the concept of the spin image [Johnson99] and then give several examples.

3.1 Our framework

Our method is divided into two stages as shown in figure 1. The first stage is completed off-line, aims to construct a visual word dictionary based on a 3D database. First, local features are extracted from each model in the database. Second, a clustering or classification method is applied to the feature collection to construct the visual word dictionary. The second stage is on-line comparison. For the query data, we extract local features and search the dictionary for the nearest visual word. We then represent the query data with a feature vector, in which each element corresponds to one visual word in the dictionary, and the value denotes the frequency of the word appearing in the query data. Finally, a certain dissimilarity metric is chosen to compare the difference between the query data and the models in the database. A retrieval rank list is the output of the framework.

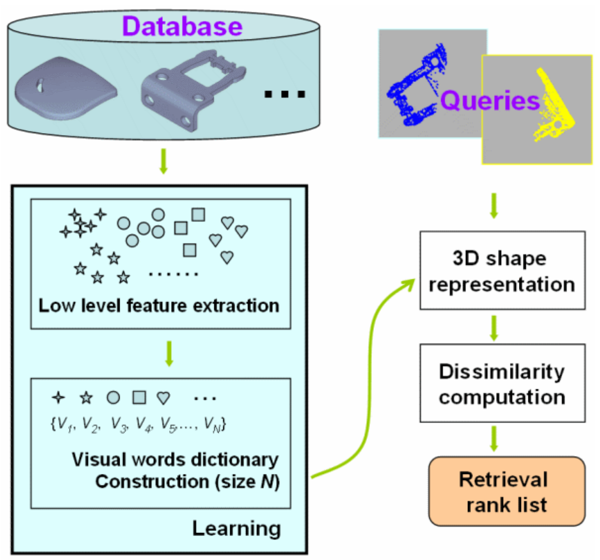


Fig. 1. Our framework

3.2 Spin image

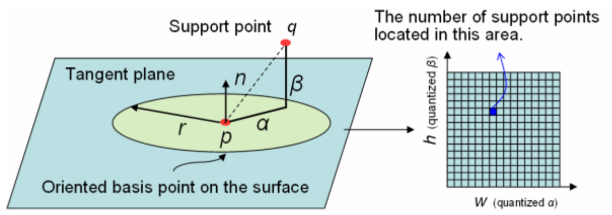


Fig. 2. Extracting low level features with spin images

As shown in Figure 2, the spin image, which is invariant to the rotation and translation transform, characterizes the local appearance properties around its basis point *p* within the support range *r*. It is a two-dimensional histogram accumulating the number of points located at the coordinate (α, β), where α and β are the lengths of the two orthogonal edges of the triangle formed by the oriented basis point *p*, whose orientation is defined by the normal *n*, and support point *q*. The final size of the spin images is defined by the width and the height of the spin plane. We choose it as the low level feature descriptor in this paper. Figure 3 demonstrates several spin images extracted from different positions from the bunny.

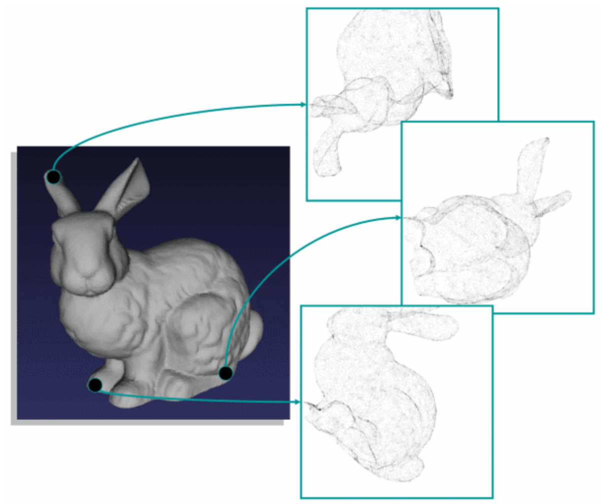


Fig. 3. Demonstration of spin images. The support range r is defined as the radius of the model. The width and height of the spin image are all 256

4. IMPLEMENTATION

In this section we will elaborate the details of the method proposed in previous section.

4.1 Low level feature extraction

Because the 3D meshes may be composed of large and tiny triangles, instead of calculating spin images based on the mesh vertices [Johnson99], a two passes sampling procedure is performed here. Using Monte-Carlo strategy [Osada02], for each

3D mesh, N_b oriented basis points p with normal n and N_s support points q are sampled uniformly on the surface in two passes respectively, where N_b=800, N_s=50000. Other parameters of spin image are defined as: 1) r =0.4R, where R is the radius of the mesh. 2) the width and height of spin images is set as w=h=10.

Now a large number of spin images are collected from the 3D shape database. Each mesh is represented with N_b spin images.

4.2 Visual words dictionary construction

With N_b*N_m spin images, where N_b is defined previously and N_m is the number of 3D meshes we used for building the visual words dictionary, k-means algorithm is applied to agglomerate N clusters. Here N equals to 1500, which defines the size of the dictionary. Therefore, each spin image is assigned with the index of its nearest cluster. Actually, other clustering algorithms [Moosmann08] can be adopted to do the work. Further research needs to be done to analyze the effects of different clustering algorithms and the size of the dictionary.

4.3 3D shape representation

For a new shape data, no matter if it is a complete model or just a partial point cloud of an object; we represent it using the visual words in the dictionary. The representation can be derived via three steps as follows:

1. Extract the low level features using spin images.
2. Calculate the distances between the spin images and the visual words. The shortest distance indicates that we can use the corresponding visual word to record this spin image.
3. Count the number of times each visual word appears on this shape.

Therefore, each shape is represented by a vector fv=(x_1, x_2, …, x_N). This is explained visually in figure 4.

Fig.4. Shape representation

4.4 Dissimilarity computation

The requirements for dissimilarity measure for the partial-to-global retrieval task are quite different than the global-to-global retrieval problem. As described in [Liu06], suppose there are query data composed of a head and a torso, it is highly probable that a human model is a candidate shape for this query. However, the human model is not a part of this query data. That means the distance between the query data and the model does not equal to the distance between the model and the query data. The dissimilarity metric should reflect this asymmetric property.

To satisfy this requirement, an ordinary symmetric distance measurement, such as L1, L2, is not a suitable choice. KL divergence is one of the metrics which satisfies the asymmetric property. We will demonstrate the different retrieval results using L1 and KL distance metric in the next section.

5. EXPERIMENTAL RESULTS

The Purdue Engineering Benchmark (PEB) [Jaynti06], which contains 801 3D CAD models, is chosen as the 3D shape database. It is classified into 42 classes such as, "Discs", "T-shaped parts" and "Bracket-like parts".

Figure 5 shows the Precision Recall curves [Shilane04] with KL divergence measurement when using different partial sizes of the object as query data. G-G means it is the PR curve for the global-to-global retrieval, P2-G means half of the original model is used as the query data, P3-G means one third of the original model is used as the query data, and so on. It verifies the intuitive feeling that less information will lead to worse retrieval results. However, even with reduced information reasonable performance is observed, suggesting robustness of the method.

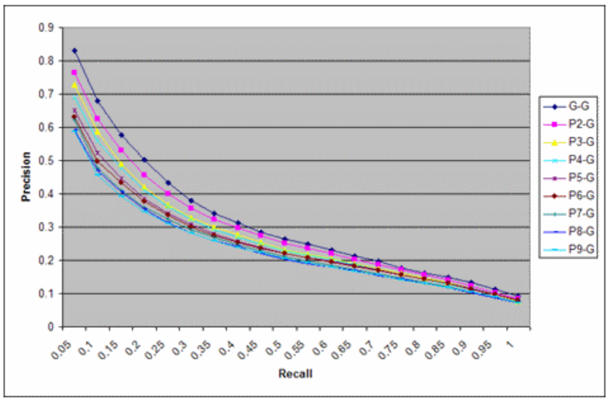


Fig. 5. The precision recall curves regarding with different size of the query data

In order to show the effects of using different distance metric, we draw two PR curves corresponding to these two metrics (see figure 6). Only one sixth of the model is used as query data.

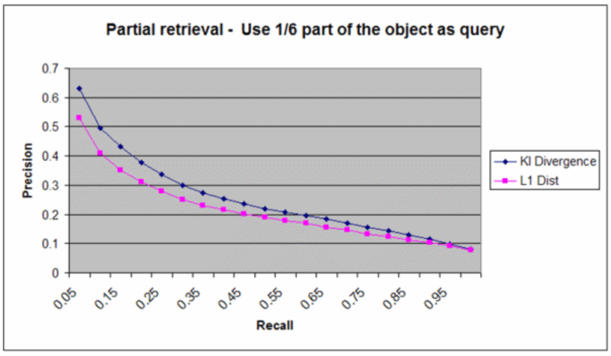
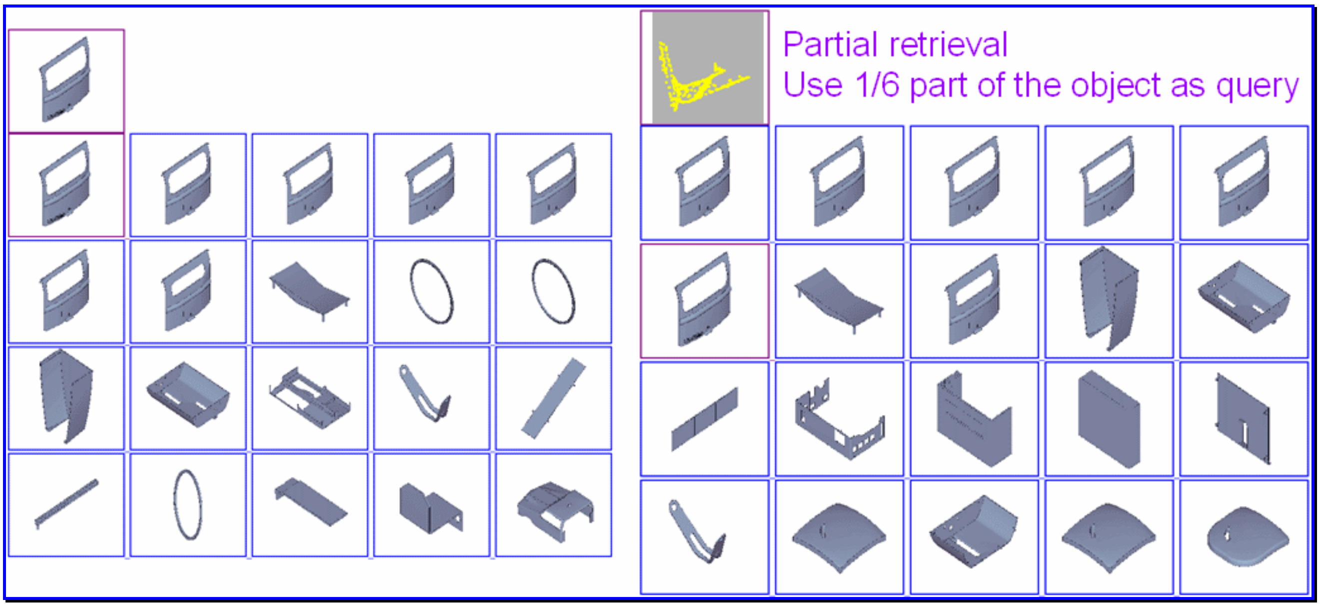
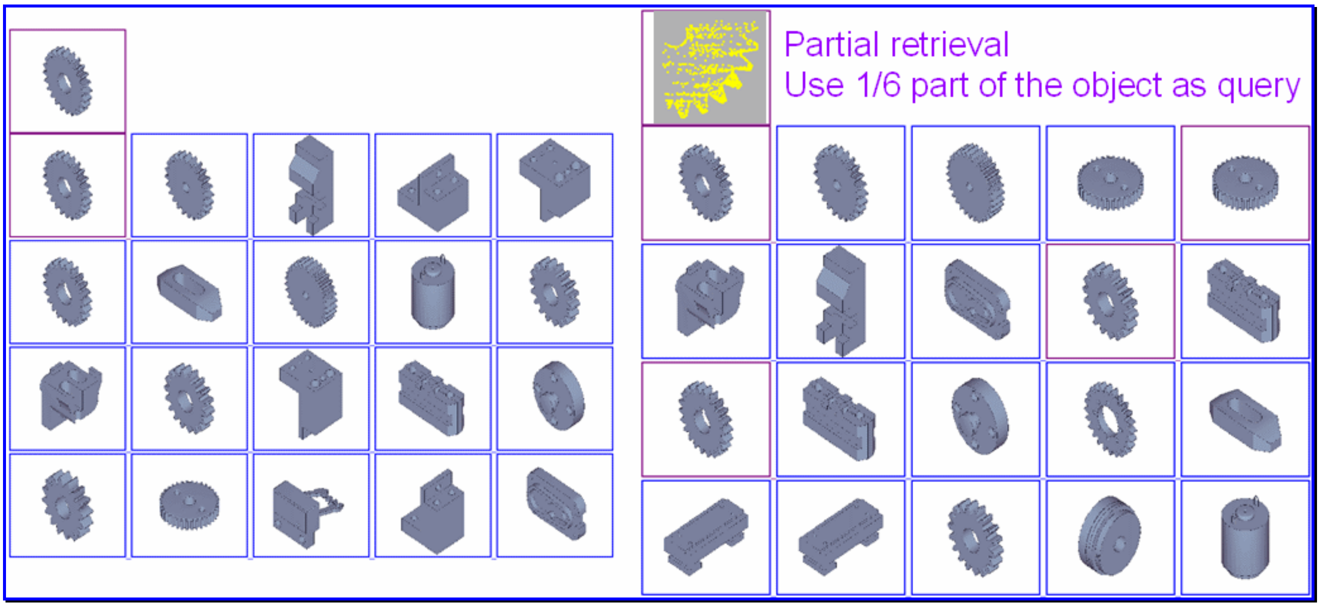


Fig. 6. The precision recall curves regarding with different dissimilarly metrics

Figure 7 provides two examples comparing the retrieval results of global-to-global retrieval and partial-to-global retrieval. The top figure shows the results when using a door as query shape. For Partial-to-Global retrieval, the left top part of the door is used as query data. In fact PEB contains only 7 door models; both G-G and P6-G retrieval rank all the 7 door models on the top of the retrieval list. The bottom figure shows results when using a gear as query shape. It shows that the P6-G retrieval is better than the G-G one, since P6-G find out more gears than G-G. Why does the partial-to-global retrieval perform better? It seems impossible. However, recalling the definition of the feature vector will provide some clues to the answer. The feature vector describes the frequency of the visual words appearing in the shape. When using the entire gear model to be the query data, the plane-kind of visual word overwhelm the other features. However, using partial of the object to be the query data, the gear teeth shape dominates the whole shape. So more gears are picked out, and listed on the top of the list.



(a) First example to show the difference between Global-to-Global (G-G) and Partial-to-Global (P-G) retrieval. The left figures show the Global-to-G-G retrieval result using a complete model (the first image listed in the first line) as the query. The right figures show the P-G retrieval result using 1/6 part of the complete model (the second image listed in the first line) as the query. The top 20 models are listed orderly according to the similarity metric.



(b) The second example to show the difference between G-G and P-G retrieval. The layout of the images is the same as that of (a).

Fig. 7. Two examples of retrieval results

6. CONCLUSIONS

In this paper, we propose to use the bag-of-words model for 3D CAD parts retrieval. The spin image is chosen as the local feature detector. We perform experiments to study the effectiveness of the method to solve the problem of partial-to-global 3D shape recognition. The results demonstrate the effectiveness of the method.

ACKNOWLEDGEMENT

We would like to thank the SIMA program and the IDUS program for supporting this work.

REFERENCES

[Chen03] D.Y. Chen, M. Ouyoung, X.P. Tian, Y.T. Shen. On visual similarity. based 3D model retrieval. Computer Graphics Forum (Eurographics '03), pp. 223-232.

[Founkhouser04] T. Funkhouser, Kazhdan, P. Shilane, P. Min, W. Kiefer, A. Tal, S. Rusinkiewicz, and D. Dobkin, Modeling by example, SIGGRAPH 2004, pp. 652-663.

[Frome04] A. Frome, D. Huber, R. Kolluri, T. Bulow, J. Malik, Recognizing objects in range data using regional point descriptors, ECCV 2004, pp.

[Ip07] C.Y. Ip, S.K. Gupta, Retrieving matching CAD models by using partial 3D point clouds, Computer-Aided Design & Applications, Vol.4, No. 5, 2007, pp. 629-638.

[Jaynti06] Jayanti, S. , Kalyanaraman, Y. , Iyer, N. , and Ramani, K. Developing An Engineering Shape Benchmark For CAD Models, Computer-Aided Design, Volume 38, Issue 9, Shape Similarity Detection and Search for CAD/CAE Applications, September 2006, Pages 939-953.

[Johson99] A.E. Johnson, M. Hebert. Using spin images for efficient multiple model recognition in cluttered 3D scenes. IEEE Trans. On PAMI, 21(5), pp. 433-449, 1999.

[Kazhdan03] Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. Symposium on Geometry Processing. June 2003.

[Li05] L. Fei-Fei, P. Perona. A Bayesian Hierarchical model for learning natural scene categories. CVPR 2005, 524-531.

[Liu06] Y. Liu, H. Zha, H. Qin, Shape topics : a compact representation and new algorithms for 3D partial shape retrieval, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'06), 2006, pp

[Mitra06] N.J. Mitra, L. Guibas, J. Giesen, M. Pauly, Probabilistic Fingerprints for Shapes, Eurographics Symposium on Geometry Processing 2006.

[Moosmann08] F. Moosmann, E. Nowak, F. Jurie. Randomized clustering forests for image classification. PAMI 2008, Vol.30, No. 9. 1632-1646.

[Ohbuchi08] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, Salient local visual features for shape-base 3D model retrieval, IEEE Inter. Conf. on Shape Modeling and Applications (SMI'08),2008, pp. 93-102.

[Osada02] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape Distributions. ACM Transactions on Graphics, 21(4):807-832, October 2002.

[Podolak06] J. Podolak, P. Shilane, A. Golovinskiy, S. Rusinkiewicz, T. Funkhouser, A planar-reflective symmetry transform for 3D shapes, SIGGRAPH 2006, pp.

[Shilane04] P. Shilane, P. Min, M.Kazhdan, T. Funkhouser. The Princeton shape benchmark. *Proc. of the Shape Modeling International 2004 (SMI'04)*, 04(00): 167-178, 2004.