Identifying Objects in Range Data Based on Similarity Transformation Invariant Shape Signatures

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
Identification and recognition of three dimensional (3D) objects in range data is a challenging problem. We propose a novel method to fulfill the task through two steps: 1) construct the feature signatures for the objects in the scene and the models in a 3D database; 2) based on the feature signature, find out the most similar model which decides the class of the corresponding object in the scene. We also evaluate the accuracy, robustness of the recognition method with several configurations. Our experimental results validate the effectiveness of our method.

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

General Terms
Algorithms, Performance, Reliability.

Keywords
Object recognition, accuracy evaluation.

1. INTRODUCTION
With the development of sensor technology, laser scanners along with digital color cameras, sonar sensors and other sensors share the role of being "perception organs" of a robot. The action of the robot is highly dependent on the information obtained from the data collected by the "perception organs". Here is a search and rescue example. When there is a gas leak in a dark factory, the robot enters as a rescuer. However, color cameras cannot capture images because of the darkness. Laser scanning becomes the preferred method for the acquisition of the information. Having obtained the range data, the robot first differentiates people from other objects, and then, carries on different strategies according to the recognition results: helping the people, and avoiding other objects.

Besides the advantage in applications, such as robot localization and strategy choice, object recognition in a cluttered scene is an interesting and challenging problem in its own right. The problem is defined as follows: given a 3D point cloud produced by a laser scanner observing a 3D scene, the goal is to identify objects in the scene by comparing them to a set of candidate objects. This is closely related to 3D shape retrieval.

The main difference between 3D shape retrieval and the recognition problem here is that, for range data, only part of the object is captured by the scanner because of the limitation of the view angle and the occlusion. The situation of occlusion is complicated. For simplicity, we only focus on the case that the occlusion does not destroy the silhouette of the objects. As a result, a complete outline is preserved, which is used as a source to construct the shape representation of the object in the scene.

For the range data, we start by segmenting it into several regions. Then each region's data is projected into a plane perpendicular to the view direction of the scanner to get a silhouette. As for the candidate objects in a 3D shape repository, their silhouettes are captured from several views. Then Fourier Mellin Transform is performed on those images to extract similarity transform invariant 2D features. After that, a comparison is done between regional range data and the models in the database to get the most similar one. The regional range data is labeled after the chosen model.

The rest of this paper is organized as follows. Section 2 describes some of the related work to our proposed approach for identifying 3D objects in range data. The procedures of feature extraction and similarity comparison are described in Section 3. In Section 4, we provide the recognition results for simulated range data and evaluate the accuracy and robustness of the approach. Section 5 provides some conclusions.

2. RELATED WORKS
There exists extensive literature addressing 3D object recognition [Bustos05]. For simple scenes, it is straightforward to use several basic geometries to represent them, such as generalized cylinders [Binford71], superquadrics [Solina90], geons [Wu94], and so on. Unfortunately, this kind of representation is too abstract to describe complicated real 3D objects. Other sophisticated methods have been put forth, including visual similarity-based [Chen03] [Vranic03], geometric similarity-based [Osada02] [Papadakis07], topologic similarity-based [Biasotti04], and local region similarity-based [Frome04]. Our approach is closely related to the visual similarity-based methods. In this section, we give a brief description about this kind of approach.

Orthogonal projected silhouette image is the most popular when researchers are thinking about using the collection of 2D images to represent the 3D shape [Chen03] [Vranic03]. Chen et al. [Chen03] first captured 100 silhouettes with 10 different configurations of cameras mounted on 10 dodecahedrons. Then 35 Zernike moments coefficients and 10 Fourier coefficients calculated from one image are concatenated as one descriptor. After that, the similarities between objects are measured using a

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

(a) A scene. **(b) The corresponding range data.** **(c) segmented scene.**

Figure 1. The scene.

particular metric. Vranic [Vranic03] only use 6 silhouettes, which is comparably less than that in [Chen03], because of the pre-alignment procedure (PCA). Actually, the recognition power of visual similarity-based methods is strongly dependent on the 2D shape descriptors and the view direction used to capture the image.

Several approaches exist to describe the shape of an image, including geometric moments, complex moments, Legendre moments, Zernike moments, Fourier descriptors, etc. [Liao96]. Fourier and Zernike, which are a contour shape descriptor and a region shape descriptor respectively, are superior to the others according to the research of Zhang et al [Zhang02]. Nevertheless, they do not completely satisfy the invariance requirement with respect to similarity transformations (i.e. rotation, translation and scale). Under the Fourier-Mellin transform framework, which is widely used in image reconstruction and image retrieval, complete invariant descriptors can be derived [Yu07].

In this paper, we investigate a new method for 3D object recognition based on orthogonally projected silhouettes under the Fourier-Mellin transform framework.

3. IDENTIFICATION PROCEDURE

The whole identification procedure is divided into 2 main parts: constructing the 3D descriptors, and computing the similarity between the range data and the candidate objects in the database. However, a scene usually includes several objects. In order to figure out what each object is, we should segment the range data.

Actually, for the range data captured from a certain view direction, it can be regarded as an image whose resolution is equal to the scanner's resolution and whose pixel values record the depth from the surface of the objects to the scanner. Thus thousands of image segmentation methods can be used [Sezgin04]. A threshold approach is applied here to segment the range data. Figure 1(a) shows a scene, while figure 1(b) is the corresponding range data, and figure 1(c) displays the segmented result in which different color refers different object.

3.1. 3D descriptor construction

For each object in the scene, after the segmentation phrase, a silhouette is obtained. Three steps of Fourier-Mellin transform are performed on the silhouette to extract a similarity invariant feature vector, which is shown in figure 2.

1) A 2D FFT

$$F(u,v)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy$$

is applied to the silhouettes.

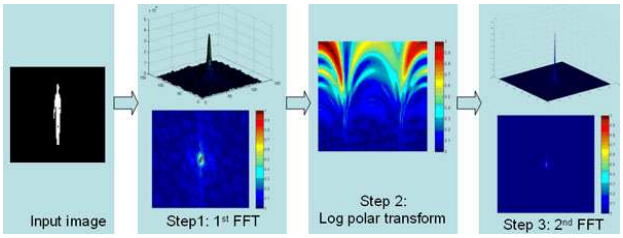


Figure 2. The procedure of FMT.

2) A log polar transform is performed on the images composed of the magnitudes of the Fourier coefficients. In this step, the resolution of the image can be changed, which defines the size of the final descriptor. The resolution is denoted as $M*M$.

3) Another 2D FFT is carried out on the log polar images to obtain the Fourier Mellin coefficients.

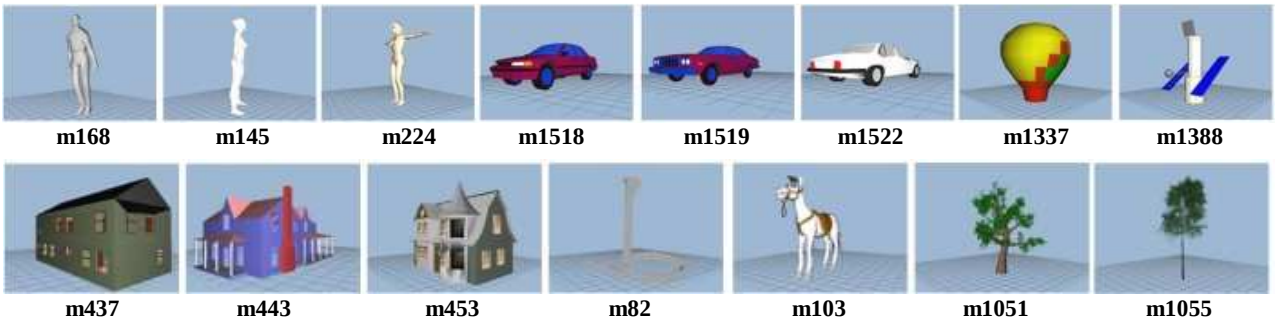
Because of the symmetry property of FFT, we choose the magnitudes of the Fourier Mellin coefficients located in the first quadrant as the descriptor, that is

$$FV=(c_1,c_2,\ldots,c_K) \quad (1)$$

where $K=M*M/4$.

To decrease the size of the feature vector, there is another solution to construct it. The coefficients in the first quadrant are summed up along x and y directions. Therefore, the form of feature vector keeps the same form as in equation (1), in which $K=M*2$.

Figure 3. The demonstration for capturing several silhouettes from defined positions on the surface of bounding sphere.



In the reference data set, m168, m145, m224 belong to class "human"; m1518, m1519, m1522 belong to class "sedan"; m437, m443, m453 belong to class "two story house"; the others belong to different classes.

Figure 4. The reference data set.

For the candidate models, the construction of descriptor is similar to that of the objects in the scene. Nevertheless, more than one such descriptor is needed to give a complete description for the model. Several cameras are placed on the surface of a sphere to fulfill this goal, whose center is the center of the model and whose radius equals to the model's max radius (shown in figure 3). The positions are defined by the longitude and latitude:

$$(\theta_i, \varphi_j), 0 \le \theta_i < 2\pi, 0 \le \varphi_j < \pi, \quad (2)$$

where $i,j=1,2,\ldots,N$. As a result, the descriptor for one candidate model is denoted as an array:

$$FA = \begin{bmatrix} f_1^1 & f_2^1 & \cdots & f_K^1 \\ f_1^2 & f_2^2 & \cdots & f_K^2 \\ \vdots & \vdots & \vdots & \vdots \\ f_1^{N*N} & f_2^{N*N} & \cdots & f_K^{N*N} \end{bmatrix} \quad (3)$$

3.2. Similarity computation

To compare descriptor *FV* eq. (1) with descriptor *FA* eq. (3), L1 distance measurement is used. The similarity is evaluated based on it:

$$sim_{l_1} = \min_j \left(\sum_{i=1}^{K} \left| f_i^j - c_i \right| \right), \quad (4)$$

where $j=1,2,\ldots,N$. The smaller the value is for equation (4), the more similar the object is to the candidate model.

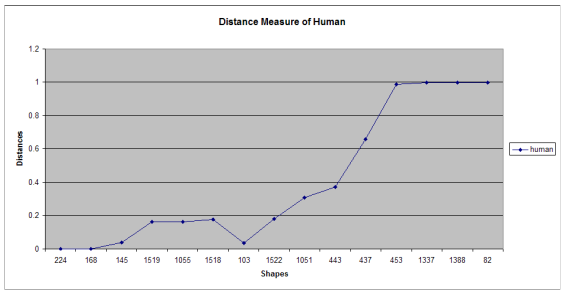
4. RESULTS AND EVALUATION

The recognition power of the method is tested on a set of reference 3D shapes made up of 15 models taken from the PSB 3D model library [Shilane04]. Figure 4 shows all of them. The point clouds used in our experiments were generated using a simulation program with resolution 256 by 256, which is regarded as query data. It is composed of one car (m1518), one person (m168) and one two-story house (m443) from the reference 3D model sets, which are all scaled to the normal size as in the real world in order to be reasonable.

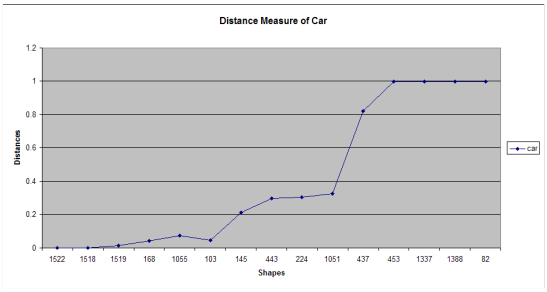
The recognition results are shown in figure 5 with $N=10$, $M=64$. The X axis represents all the reference models in the data set, while the Y axis indicates the distance between the query object and the models.

How the noise level affects the recognition results is discussed here. Take one object –house – in the query scene as an example, we add Gaussian noise to it and get the result with different noise levels. The recognition curves for different noise level are shown in figure 5. From the plot, we know that the noise will change the distance value a little bit, but the appearance of the curve keeps similar to the original one (the dark blue curve), which shows the robustness of our method.

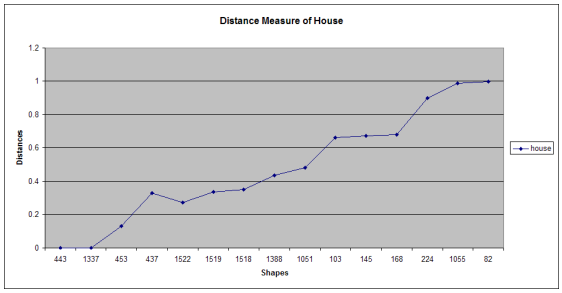
Since the range data can be extracted from different directions, the robustness of the algorithm related to the extracting directions should be evaluated. The 3D range scanner is placed on the surface of the bounding sphere of the scene. And the positions are defined by the longitude and latitude (eq. 2) with $N=15$, which means the amount of the range data from different directions is 225. The configuration guaranteed the differences between the shooting direction of the camera and the scanning direction of the range scanner. Taking the model of a person (m168) as an example, figure 6 shows the effect of the scanning directions, in which the x axis shows the name of the model and the y axis shows how many times the object is recognized as the model. It shows the accuracy is 87%.



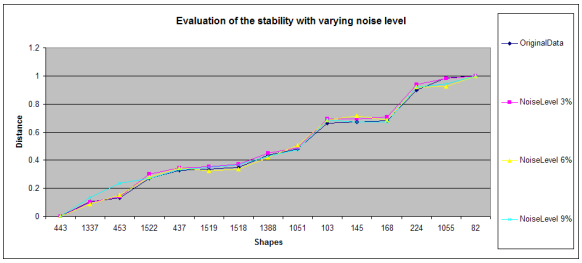
(a) Recognition result for "human".



(b) Recognition result for "Car".



(c) Recognition result for "House".



(d) Recognition result with different noise level for "House".

Figure 5. The recognition results (a)(b)(c) for all the objects in the scene. The X axis records the shapes from the reference set. (d) represents the recognition results for "House" with different noise level.

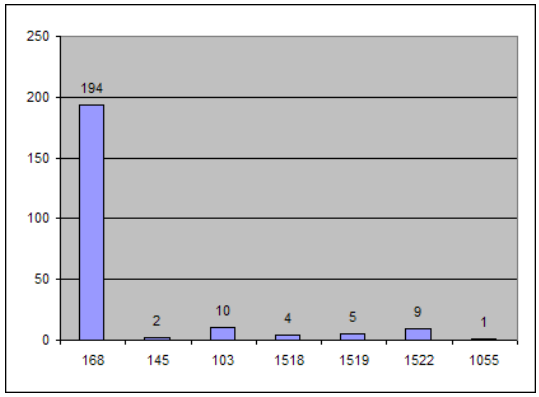


Figure 6. The effect of the shooting directions.

5. CONCLUSION

In this paper, we presented a novel procedure for 3D object recognition using the Fourier-Mellin transform. Although FMT is widely used in 2D image retrieval and reconstruction, it has not been used in 3D shape recognition. We have applied it to 3D shape recognition, and the experimental results show its effectiveness. Furthermore, how the noise level and the scanning direction affect the recognition result is investigated. Nevertheless, to guarantee the recognition accuracy, the completeness of the silhouette should be kept, which is rare in a cluttered and clustered scene. In future work, more stable local feature signatures will be introduced to alleviate the effect of cluttering and clustering.

ACKNOWLEDGEMENT

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

REFERENCES

[Biasotti04] S. Biasotti. Reeb graph representation of surfaces with boundary. Proc. of the Shape Modeling International 2004 (SMI'04), 2004.

[Binford71] T.O. Binford. Visual perception by computer. IEEE Conf. on Systems and Control, Miami, FL, 1971.

[Bustos05] B. Bustos, D. A. Keim, D. Saupe, T. Schreck, D. V. Vranic. Feature-based similarity search in 3D object databases. ACM Computing Surveys (CSUR), 37(4):345-387, 2005.

[Chen03] D.Y. Chen, M. Ouhyoung, X.P. Tian, Y.T. Shen. On visual similarity based 3D model retrieval. Computer Graphics Forum (Eurographics '03), 03: 223-232, 2003.

[Frome04] A. Frome, D. Huber, R. Kolluri, T. Bulow, J. Malik. Recognizing objects in range data using regional point descriptors. Proc. of European Conference on Computer Vision (ECCV), Prague, Czech Republic, 2004.

[Liao96] S. X. Liao, M. Pawlak. On image analysis by moments. IEEE Trans. Pattern Analysis and Machine Intelligent, 18(1996): 254-266, 1996.

[Osada02] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin. Shape distributions. ACM Transaction on Graphics, 21(4):807-832, 2002.

[Papadakis07] P. Papadakis, I. Pratikakis, S. Perantonis, T. Theoharis. Efficient 3D shape matching and retrieval using a concrete radicalized spherical projection representation. Pattern Recognition, 40(2007):2437-2452, 2007.

[Sezgin04] M. Sezgin, B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging, 13(1), 146-165, Jan. 2004.

[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.

[Solina90] F. Solina, R. Bajcsy. Recovery of parametric models from range images: the case for superquadrics with global deformations. IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI), Feb. 1990.

[Vranic03] D.V. Vranic. 3D model retrieval. Ph. D. Dissertation, 2003.

[Wu94] K. Wu, M. Levine, Recovering parametric geons from multiview range data. CVPR, June, 1994.

[Yu07] H. Yu, M. Bennamoun. Complete invariants for robust face recognition. Pattern Recognition, 40(2007):1579-1591, 2007.

[Zhang02] D. S. Zhang, G. Lu. An integrated approach to shape based image retrieval. Proc. of 5th Asian Conf. on Computer Vision (ACCV), 02: 652-657, January 2002.