Shape Representation and Matching of 3D Objects for Computer Vision Applications

YASSER EBRAHIM[1], WEGDAN ABDELSALAM[1], MAHER AHMED[2], SIU-CHEUNG CHAU[2]
[1]Dept. of Computing & Information Science
University of Guelph,
Guelph, Ontario, N1G 2W1
CANADA

[2]Dept. of Physics and Computer Science
Wilfrid Laurier University,
Waterloo, Ontario, N2L 3C5
CANADA

Abstract: -In this paper we present a novel approach to 3D shape representation and matching based on the combination of the Hilbert space filling curve and Wavelet analysis. Our objective is to introduce a robust technique that capitalizes on the localization-preserving nature of the Hilbert space filling curve and the approximation capabilities of the Wavelet transform. Our technique produces a concise 1D representation of the image that can be used to search an image database for a match. The technique exhibits robustness in cases of partial and occluded image matching. Our technique is translation, scale, and stretching invariant. It is also robust to rotation around the Y axis.

Key-Words: -3D Shape representation, Shape matching, Hilbert curve, Wavelet transform, Grey level

1 Introduction

Object recognition is a key part of any robotic vision system. The process of recognition is one of the hardest problems in computer vision, however. Although human can perform object recognition effortlessly and instantaneously, an algorithmic description of this task for implementation on machine has been very difficult especially in case of 3-D objects. One popular approach to 3D object identification is to model 3D objects and compare a search object to these models. The advantage of this approach is the ability to recognize objects from the image of a single view [1][2]. However, a single view may not contain sufficient features to recognize the object. In addition, it requires complex feature sets making the recognition process time consuming [3]. To overcome this problem, modeling 3D object recognition using multiple 2D views was proposed as an alternative. Edelman and Bulthoff [4] found a strong and stable correlation between recognition performance and viewpoint variation and suggested object representations by multiple viewpoint 2D representations. Murase and Nayar [5] and Nene [6] developed a parametric eigenspace method to recognize 3D objects directly from their appearance. This technique however is not robust to occlusion and do not provide indication on how to optimize the size of the database with respect to the types of objects considered for recognition and their respective eigenspace dimensionality.

In this paper we present a robust 3D object representation and matching technique that minimizes the number of 2D images needed to represent a 3D object without sacrificing much of the retrieval capabilities of the system.

Our approach is based on the use of the Hilbert space filling curve to produce a concise 1D representation of each object. Wavelet approximation is used to smooth out noise and fine details while keeping the main features intact.

1.1 The Hilbert space filling curve

A space filling curve is a continuous path, which visits every point in a 2-dimensional grid exactly once and never crosses itself. Space-filling curves provide a linear order of the points of a grid. The goal is to do so while keeping the points that are close in 2-dimensional space close together in the linear order [7]. The Hilbert curve is one of the most popular space filling curves in use because of its excellent localization capabilities. Fig.1 shows a Hilbert curve at 3 different levels of detail (adopted from [7]).

The basic Hilbert curve is denoted by *H1*. Curve *H1* has four vertices at the center of each quarter of the unit square. Curve *H2* has 16 vertices each at the center of a sixteenth of the unit square. To derive a curve of order $i+1$, connect four copies of *Hi*, each rotated as necessary [7].
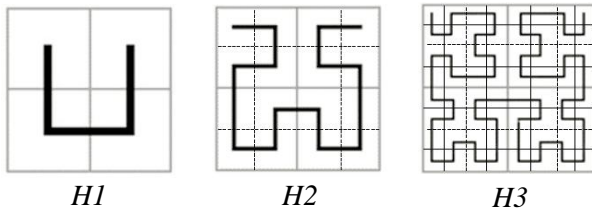
H1   *H2*   *H3*

Fig.1. Hilbert curves of order 1, 2, and 3.

1.2 The Wavelet transform

The aim of discrete wavelet analysis is to represent a function using a sum of details of smaller and smaller scales. So, it should be possible to distinguish local information from global information [8]. In this paper we focus on the global features of a shape represented in the approximation vector produced by the 1D Wavelet transform.

This paper is organized as follows; in section 2 we describe our proposed technique. In section 3, we list the different advantages of our approach with some examples. In section 4 we present the results of an experiment conducted.

2 Our approach

Our approach to 3D shape representation and matching can be summarized as follows.

2.1 Building the object database

To build the object database we extract the 3D shape of each object by extracting the 2D shape of the object from different view points. For this puropse we need 5 different images representing the object from angles 0, 45, 90, 225, and 270. The first three are used for the frontal view of the object and the last two are used for the rear view of the object. Each of these images is processed as follows:

- Find the object's minimum bounding rectangle and crop the image accordingly to ensure translation invariance
- Scan the image based on the Hilbert filling curve saving the gray level of each pixel scanned to a vector V
- Apply the Wavelet transform to V and sample the result to produce the *shape vector, v.*

See Fig.2 for an example of these steps. Notice that the shape curves for angles 135, 180, and 315 can be derived from the existing shape curves as they are mirror images of existing images as can be seen in Fig.3 (we assume that objects are symmetric). Because of the U shaped Hilbert curve we use to scan the image (see Fig.1), the shape vector of an existing image can be reversed to approximate the shape vector of its mirror image. See Fig.4 for an example.

2.2 Searching the database for a match

- Apply the shape extraction steps above to the search object
- Compare the shape vector for the search object to each of the five shape vectors representing each object in the database computing the correlation between them. The highest correlation of the five represents how close the search object is to the database object at hand.
- The database object(s) with the highest correlation are returned as possible matches.

Step	Example
Read image	
Find the minimum bounding rectangle and crop	
Use the Hilbert curve to scan the image and produce V	
Apply the Wavelet transform and sample the result to produce v	

Fig.2. The basic steps of our approach

3 Advantages of our approach

We believe that the strongest points of our approach is its ability to address a number of diverse issues in 3D shape representation and matching that are very rarely addressed collectively by any one approach.

3.1 Invariance to translation, scaling and stretching.

We achieve translation invariance by scanning the minimum bounding rectangle of the object rather than the full image which might have blank spaces around the object. Scale invariance is achieved by the virtue of the fact that the matching is dependent on the shape of the resulting shape vector curve rather than its values. Because scanning the image follows the same path

regardless of its size, similar objects produce similar curves. The technique is also invariant to scaling in one dimension (stretching). This is because the Hilbert curve stretches to fill the area regardless of its shape. It works equally well with square and rectangular areas (see Fig.5 for an example).
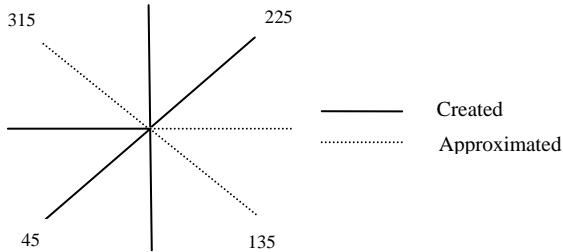

Fig.3. A 5 shape curve representation of an image. Each line represents a shape curve representing the object from a certain angle.

Fig.4. Reversing a shape curve to approximate that of the mirror image

3.2 Robustness to occlusion

Because we use the shape of the entire curve to find the correlation between two shapes, similar objects that differ in a small area due to occlusion usually still exhibit a fairly strong correlation. In many cases it is possible to pinpoint the occluded part of the image by finding the part of the curve with the lowest correlation value. See Fig.6 for an example.

3.3 Ability to perform partial searches

Because the shape vector actually reflects the localization of the different parts of the object, it is fairly easy to pinpoint which part of the 2D object is represented by which segment(s) of the shape vector. To conduct a partial search, the user needs to provide a part of an object and its approximate location on the full object. The system will then try to find a match by correlating the shape vector for the segment provided with the corresponding part of the shape vector for each of the objects in the database. See Fig.7 for an example

of how each portion of the image can be mapped to a portion of the shape curve.

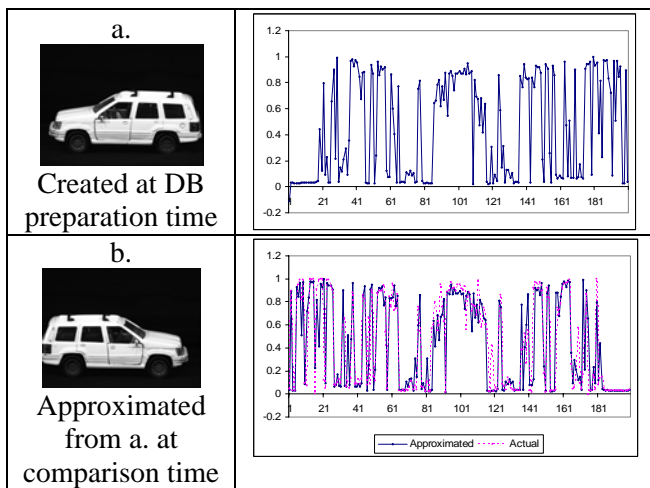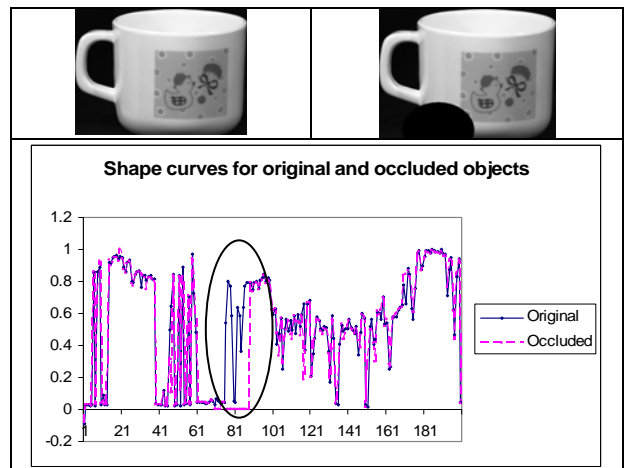Fig.5. The shape curve for an image and a stretched version of it.

Fig.6. The shape curve of an image and an occluded version of it. Notice how the occlusion is localized on the curve allowing us to pinpoint its location on the original image.
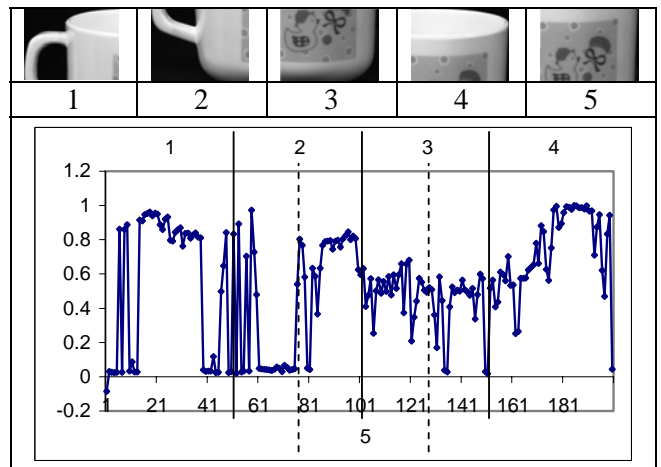

Fig.7. Five different parts of the image and their mapping to the shape curve

3.4 Capturing of color, texture, and boundary information

Because our technique captures the grey level of each pixel, the resulting shape curve actually captures more than just the shape. It captures the grey level of each region of the image as well as its texture. This means that matching objects using their shape curve actually involves matching them based on shape, texture and color.

3.5 Robustness to vertical rotation

This feature is particularly important since vertical rotation could cause major identification difficulties due to the 3D nature of objects surrounding a robot and the 2D nature of images captured by its camera. Our suggested representation exhibits a high level of robustness to vertical rotation as can be seen in Fig.8. From fig.8 we can see that the two shape curves are highly correlated which means that the vertical rotation didn't affect the representation in a significant way. As a result, very few shape curves are actually needed to represent the object from different camera points of view. As we will show in section 4, in most cases, only five shape curves are needed to represent an object.
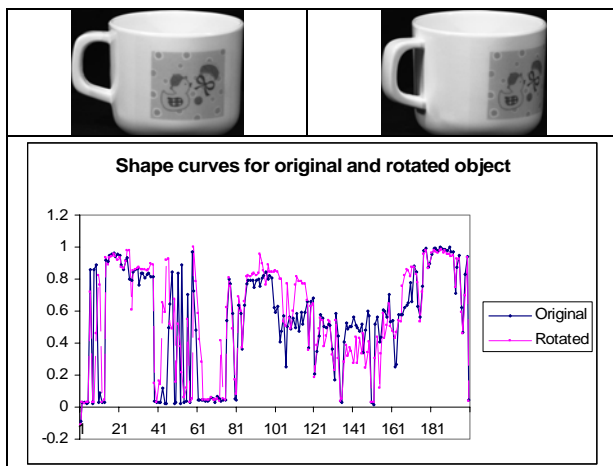

Fig.8. The shape curve for an object and the same object rotated by 30 degrees

4 Experimental results

To test the efficacy of our approach we experimented with five image groups obtained from the Amsterdam Library of Object Images (ALOI) [9]. See Fig.9 for samples of the five groups we used for our experiment.

For each group we used a set of images depicting the objects from angles 0, 5, 10,…, 80, and 85. We calculated the correlation between the shape curve of each image and that of the image depicting the object rotated by 5, 10, 20, 25, and 30 degrees. For example, in table 1, the correlation 0.79 on the last column is the

correlation between Mug-5 (the mug rotated 5 degrees) and Mug-35. The average correlation within each group for each of these categories was calculated. Fig.10 shows the results for the five groups.
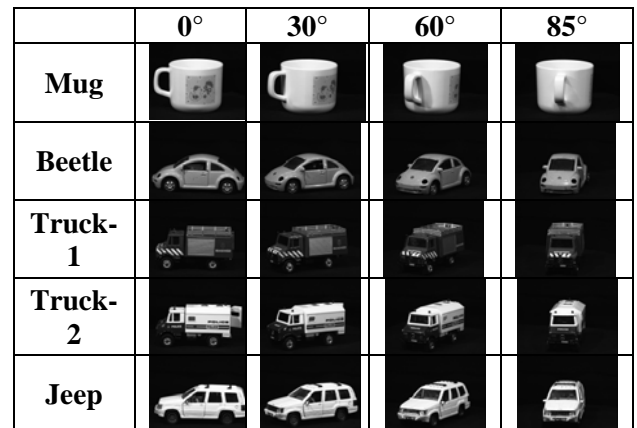

Fig.9. Samples of the 5 groups of images used for the experience.

Table 1. The correlation between each image and the image rotated 5, 10, 15, 20, 25, and 30 degrees.

	5° diff.	10° diff.	15° diff.	20° diff.	25° diff.	30° diff.
Mug-0	0.95	0.92	0.90	0.86	0.84	0.82
Mug-5	0.97	0.94	0.90	0.90	0.84	0.79
Mug-10	0.97	0.93	0.93	0.87	0.82	0.80
Mug-15	0.96	0.95	0.91	0.86	0.84	0.77
Mug-20	0.96	0.91	0.87	0.83	0.78	0.70
Mug-25	0.92	0.86	0.84	0.76	0.71	0.60
Mug-30	0.97	0.89	0.83	0.75	0.68	0.57
Mug-35	0.91	0.87	0.78	0.73	0.61	0.52
Mug-40	0.93	0.87	0.77	0.66	0.58	0.54
Mug-45	0.93	0.87	0.77	0.68	0.61	0.60
Mug-50	0.89	0.80	0.71	0.64	0.64	0.64
Mug-55	0.92	0.84	0.74	0.72	0.72	0.70
Mug-60	0.93	0.83	0.78	0.77	0.76	
Mug-65	0.93	0.84	0.81	0.77		
Mug-70	0.90	0.87	0.82			
Mug-75	0.96	0.89				
Mug-80	0.92					
Average	**0.94**	**0.88**	**0.82**	**0.77**	**0.73**	**0.67**

From Fig.10 we can see that the correlation did decrease almost linearly as the rotation angle increased. The average correlation decrease was about 5% for each 5 degree rotational difference. This means that if we use the suggested five shape curves to represent the image from all angles (after approximating mirror images as explained above), the average maximum correlation reduction due to vertical rotation will be equal to 45/2/5 * 5% = 22.5% where the 45 is the degree difference between any two consecutive images. We divide the angle (i.e., 45) by two because for an image I_j, 45/2=22.5° is the largest angle the image could be away

from two consecutive images I_i and I_{i+1} where the angle $A(I_i) \leq A(I_j) \leq A(I_{i+1})$. Adding two more shape curves (see Fig.11) can reduce the average correlation reduction to 30/2/5 * 5% = 15%.
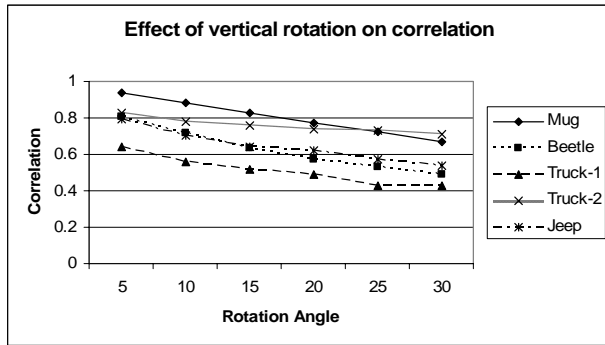

Fig.10. The effect of vertical rotation angle on correlation

Because different images may exhibit different degrees of correlation degradation due to vertical rotation, an adaptive version of our technique can be easily implemented such that the number of shape curves representing an image is determined based on both the severity of, and degree of tolerance to, correlation degradation.
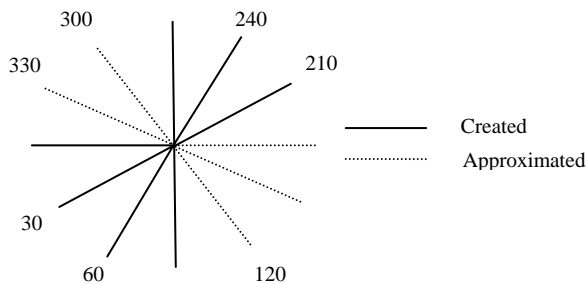

Fig.11. A 7 shape curve representation of an image at 30 degree intervals

5 Conclusion and future work

In this paper we presented a novel approach to 3D shape matching and representation using the Hilbert space filling curve and the Wavelet transform. The localization feature of the Hilbert filling curve allows this approach to tackle many tough problems with shape matching such as occlusion and vertical rotation. We also showed how the technique is robust to translation, scaling, and stretching.

In spite of the promising results we got using this technique; there is still room for improvement. We are currently investigating the possibility of using the most distinctive portion(s) of the shape vectors for correlation calculation rather than the whole shape vector. The idea is to identify the *key* part(s) of the shape vector that can be used to get the best matching results given the current database. When comparing the database object to a search object, only the key part(s) of the shape vector will be compared to the corresponding part(s) in the shape vector of the search object. Besides minimizing search time, we believe this technique will results in more, higher-ranked correct hits.

We are also investigating the use of the detail information provided by the Wavelet transform to improve the retrieval rate by reducing false positives.

References:
[1] Besl, P. J. and Jain, A. C. Three-dimensional Object Recognition, *ACM Computer Survey.* 17:76-145.1985.
[2] Elsen, I.; Kraiss, K. –F.; and Krumbeigel, D. Pixel Based 3D Object Recognition with Bidirectional Associative Memories. *International Conference on Neural Netwoks.* 3:1679-1684, 1997.
[3] Ham, Y. K. and Park, R. –H. 3D Object Recognition In Range Images Using Hidden Markov Models And Neural Networks. *Pattern Recognition.* 32:729-742,1999.
[4] Edelman, S. Y.; Bulthoff, H. H.; Tarr, M. J., How Are Three-Dimensional Objects Represented In The Brain?. *Technical Report* , MIT. 1994.
[5] Murase, H. and Nayar, S. K, Visual Learning And Recognition Of 3d Objects From Appearance. *International Journal of Computer Vision.* 14:5-24. 1995.
[6] Murase, H., Nayar, S. K, and Nene, S. A. Real-Time Object Recognition System. *Proc. IEEE International Conference on Robotic and Automation.* 1996.
[7] Wegdan Abdelsalam, Jay Black, Maintaining quality of service for adaptive mobile map clients. *Master thesis*, University of Waterloo, 2002.
[8] C.-H. Lamarque, F. Robert, Image analysis using space-filling curves and 1D wavelet bases, *Pattern Recognition*, Volume: 29, Issue: 8, 1996.
[9] MSTERDAM LIBRARY OF OBJECT IMAGES (ALOI) , http://staff.science.uva.nl/~aloi/
[10] Z. Tu and A. L. Yuille, Shape Matching and Recognition – Using Generative Models and Informative Features, *In proceedings of ECCV*, 2004.
[11] T. Adamek, N. E. O'Connor, A Multiscale Representation Method for Non-rigid Shapes With a Single Closed Contour, *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 14, No. 5, May 2004.
[12] I. El-Rube, M. Kamel, and M. Ahmed, Wavelet-Based Affine Invariant Shape Matching and Classification, *ICIP International Conferences on Image Processing*, October 2004, Singapore.
[13] E. A. El-Kwae, M. R. Kabuka, Binary object representation and recognition using the Hilbert morphological skeleton transform, *Pattern Recognition* 33(10): 1621-1636 (2000).