Segmentation And Recognition Of 3D Objects In Automatic Navigation *

Tiantian Huang, Huimin Ma, Fengting Li
Department of Electrical Engineering, 3D Image Simulation Laboratory
Tsinghua University, Beijing, 100084, China
htt05@mails.tsinghua.edu.cn
mhmpub@tsinghua.edu.cn
lift@tsinghua.edu.cn

Abstract

This paper presents a segmentation and recognition method of 3D objects in automatic navigation. In the first part of this paper, a segmentation method based on nautical scene is proposed, which is composed of image preprocess, ROI detection, subarea process and object detection. In addition, an applied visual resolution calculation method is presented to control the simplification of original 3D model. Then a novel clustering method is introduced to merge the aspect graphs ulteriorly after viewpoint space partition to the simplified model. Finally, the segmented object will be sent to compare with the aspect graphs of the 3D model, using an improved method of Fourier descriptor. A set of experiments based on ship models are designed and implemented, and the results demonstrate the effectiveness of the method proposed in this paper.

1. Introduction

3D object recognition is now widely used in the areas of automatic navigation. There are two primary factors should be considered in an applied navigation system: one is the Real-time scene segmentation, and the other one is the gradual recognition for the image sequence.

Some research works related to the 3D object segmentation and recognition in automatic navigation have been already presented. Current nautical scene segmentation methods include threshold-based segmentation to the detected ROI [1] and contour detection by an edge-histogram approach [2]. These segmentation methods haven't considered the influence of ocean wave and other coast background. Therefore they could be applied to the images with simple background such as proportional surface, but not suitable for the images with complicated nautical background. The primary recognition methods in automatic navigation include k-nearest neighbor [1], neural network [2] and principle component analysis [3]. Most of the recognition methods, as mentioned above, need training by large numbers of samples. In addition, these methods haven't considered the influence of human visual mechanism to the gradual recognition.

This paper focuses on two issues: the Real-time scene segmentation and the gradual recognition. Considering about the complication and segmentation effect, a novel segmentation method applicable to images of the nautical scene is first introduced in this paper. Then the visual resolution could be calculated by a proposed visual equation, which will be used to control the 3D model simplification [4]. Since the observation condition changes, aspect graphs at different resolution will be generated by the simplified model, which lays on a foundation of the gradual recognition.

The paper is organized as follows: Section 2 describes the entire system modeling process. Section 3 proposed a novel segmentation method. Section 4 gives an applied visual equation. Section 5 introduces a new method of clustering. Section 6 introduces the matching strategy of 3D recognition. Section 7 presents the experiments and results of the presented method. Finally section 8 concludes the paper.

2. System modeling

The entire recognition system is mainly composed of two parts, the 3D model library foundation and the object recognition. In the library foundation part, the original 3D model will be simplified according to the result of the visual equation. Then the viewpoint space partition [5] will be applied to each resolution of the simplified models. The multiresolution aspect graph representations of the 3D model could be achieved after an additional clustering operation

*This work is supported by the National Natural Science Foundation of China (No. 60502013), and by 863 project(No.2006AA01Z115).

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.349

to the partition results. In the object recognition part, target object will be segmented from the 2D real scene. Then it will be sent to compare with the aspect graphs in the model library.

In the following parts of this paper, the methods of image segmentation, visual equation, clustering and recognition will be mainly introduced.

3. Segmentation

Image segmentation is a very pivotal step for the object recognition. Current methods in image segmentation include threshold segmentation, edge detection and statistical methods. As illustrated in Figure 2(a), there exist stochastic noise, surface waves and ocean waves in the real scene image, which could aggravate the difficulty of segmentation. Combining the features of the nautical scene, the segmentation method in this paper could be described as follow steps:

1. Preprocess the image to remove noises and enhance the image contrast, using morphological filter [6] and histogram equalization [7].

2. Segment the enhanced image by Otsu method [8], and extract the minimum bounded rectangle of the object area, which will be called ROI.

3. Segment the ROI of step 2 by Otsu method once more.

4. Reprocess the segmented area of step 3 if it is larger than a threshold. First detect the edge of the area, and then extract the longest straight line by Hough transform [9], which could be the coastline in most situations. According to the coastline, the coast background the object could be successfully separated.

5. Identify the objects by the features of area and length breadth ratio, and remove the fake objects.

(a) (b) (c)

Figure 1. Contrast of the segmentation results. (a) is the original image. (b) is the segmentation result of the method [1]. (c) is the segmentation result of the method proposed in this paper.

Illustrated in Figure 1, the segmentation result of the method proposed in this paper keeps more details of the image contour, which is benefit to the object recognition.

4. A visual equation

According to the resolution limit theory [10], the resolution of human eyes is limited. Therefore it's significative to find a 3D model representation method based on human visual mechanism. Most of contemporary 3D model representation methods haven't considered the observation condition because no applied imaging system model has been established, and visual resolution under different conditions is hard to calculate.

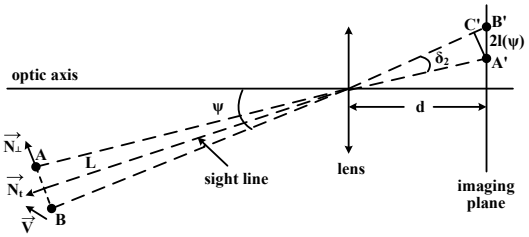


Figure 2. The imaging system model.

A new imaging system model based on receptor theory [10] is illustrated in Figure 2. We assume that the distance between the object and the lens is much larger than the object size, which is true in most situations. Therefore the distances between the lens and all the points of the object can be considered the same, and the similar for the approximation of the angle ψ in Figure 2.

Assuming that L is the distance between the object and the lens, d is the distance between the lens and the imaging plane, and ψ is the departure angle between the object and the optic axis. Denote $\overrightarrow{V}$ as the velocity of object and T as the exposal time. Let $\overrightarrow{N_t}$ denote the unit vector parallel with the connection line between the object and the lens, $\overrightarrow{N_\perp}$ denote the unit vector parallel with $\overrightarrow{BA}$ and vertical with $\overrightarrow{N_t}$. Define $l(\psi)$ as the scale of cone cells at the angle ψ.

According to the receptor theory, human eyes could only distinguish two object points when there exist non-stimulated cone cells between two stimulated cone cells. Therefore the minimum distance of two distinguished imaging points could be expressed as follows:

$$\|\overrightarrow{A'B'}\| = 2l(\psi). \tag{1}$$

If the object doesn't move in the direction of $\overrightarrow{N_\perp}$, the visual resolution could be calculated as follows:

$$\Delta S = \|\overrightarrow{AB}\| = 2l(\psi)L\cos^2\psi/d. \tag{2}$$

In addition, considering about the exposal time, tracks will be left in the imaging system by points of the high speed object during the exposal time. With consideration of this circumstances, the visual resolution should be modified as follows:

$$\Delta S(L, \psi, \overrightarrow{V}) = 2l(\psi)L\cos^2\psi/d + \|\overrightarrow{V} \cdot \overrightarrow{N_\perp}\|T. \quad (3)$$

According to Eq. (3), the visual resolution under current condition could be calculated, which will be used to control the 3D model simplification. As the observation condition changes, simplified models at different resolution could be obtained. Therefore the object could be successfully expressed by a set of multi-resolution aspect graphs generated by the simplified models.

5. Clustering

Considering about the aspect graphs with close viewpoints might have little difference, and could not be distinguished in the recognition method. To reduce the insignificant matching cost, similar aspect graphs could be merged by an additional clustering process. A novel clustering method based on a tree structure is proposed in this paper, which describes the differences between aspect graphs at multiple hierarchy. Fourier descriptor is used for the similarity measurement between single aspect graphs, which will be introduced in detail in section 6.

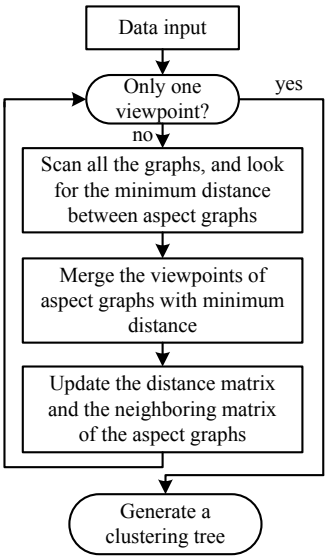


Figure 3. The clustering tree foundation.

Figure 3 illustrates the foundation process of a clustering tree. The viewpoints of similar aspect graphs will be merged step by step until all the partitioned viewpoints have been merged into one viewpoint space. As shown in Figure 4, the clustering result of each step will be presented in a clustering tree.

The original aspect graph space has the highest resolution, and discretional two aspect graphs could be distinguished. As the requirement of resolution changes, the corresponding threshold could be selected to control the degree of aspect graph merging. Illustrated in Figure 4, the horizontal axes shows the number of the aspect graph, and the vertical axes shows the distance between aspect graphs. Draw a horizontal line which intersects the vertical axes at a point with a value of the threshold. The times of the intersection between the horizontal line and the clustering tree shows the number of aspect graphs after merging, for example, 4 aspect graphs have been left in Figure 4. Since clustering is carried out from bottom to top, and no aspect graphs have intersection, the clustering tree after truncation could still be consider as a normal tree. The final result of clustering includes all the aspect graphs above the cutting line, and one representable aspect graph of each aspect graph set under the cutting line.

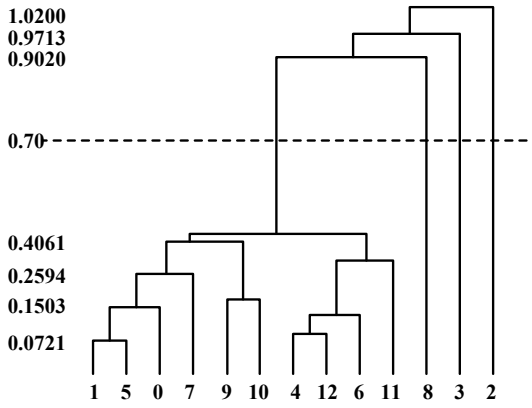


Figure 4. The clustering tree truncation.

6. Matching and recognition

The similarity measurement between aspect graphs is essential for both clustering and matching processes. There are two primary features applied during the 3D object recognition process: one is the shape-based description [11], and the other one is topology description [12]. Considering about the high speed, easy storage, and low features dimension, the shape-based description has been more valuable during practical application. Fourier descriptor [13], one of the shape-based description, will be introduced in this paper for its fast calculation and the robustness to noise and other boundary deformations.

In the Ordinary Fourier descriptor method [13], the boundary points were directly used for the input spatial sequence. Therefore the Fourier coefficients will be related

with the shape scale, sampling direction and initial position of the image contour. To eliminate above infection, an improved Fourier descriptor method is introduced into this paper, which uses the center distance for the input spatial sequence, and normalizes the Fourier coefficients. The recognition method based on Fourier descriptor could be described as follows:

1. Establish a coordinate system for the 2D images. The object centroid could be calculated as follows:

$$x_c = \frac{1}{N}\sum_{k=0}^{N-1} x_k, y_c = \frac{1}{N}\sum_{k=0}^{N-1} y_k. \quad (4)$$

where N is the number of points of the object contour.

2. Calculate the center distance of all boundary points as follows:

$$u_t = [(x_k - x_c)^2 + (y_k - y_c)^2]^{1/2}, t = 0, 1, ..., N-1. \quad (5)$$

3. Apply the Fourier transform to the center distance sequence calculated by step 2, and the Fourier coefficients could be expressed as follows:

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} u_t e^{-j2\pi nt/N}, n = 0, 1, ..., N-1. \quad (6)$$

4. Normalize the Fourier coefficients as follows:

$$f = [\frac{\|a_1\|}{\|a_0\|}, \frac{\|a_2\|}{\|a_0\|}, ..., \frac{\|a_{N-1}\|}{\|a_0\|}]. \quad (7)$$

5. Assume that a, b are the number of boundary points of image i and image j. The distance between image i and j could be calculated as follows:

$$d(i,j) = \sum_{k=1}^{min(a,b)} (f_{ik} - f_{jk}). \quad (8)$$

6. Denote s as a 2D segmented object and M as a 3D model. Assume that n is the number of aspect graphs in the library of model M, and M_j is one of the aspect graphs of model M. The distance between the segmented object s and the 3D model M could be calculated as follows:

$$D(s, M) = min(d(s, M_j)), j = 1, 2, ..., n \quad (9)$$

7. Define T as the model sets and m as the number of models. Current object will be recognised as model I, if $D(s, I) = min(D(s, T_i)), I \in T, i = 1, 2, ..., m$.

7. Experiments and results

All the 3D models used in this paper are obtained from the standard 3D model library published by Princeton University [14]. Five types of models from this library are selected as test objects, including two types of ships, two types of planes and one type of car. In our simulation system, the original 3D ship model will be conversed to the 2D image after a series processes of OpenGL texture mapping [15], gauss filtering and image enhancement. The experiments consist of two parts: aspect graph library foundation and 3D object recognition.

7.1. Aspect graph library foundation

The parameters set by the simulation system are listed in Table 1.

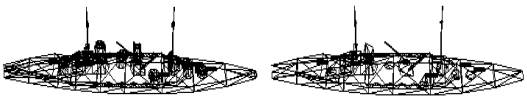
Table 1. Parameters of the simulation system

CCD pixel	CCD width	CCD height	Lens focus
7 million pixels	5.76mm	4.29mm	5.8mm
L	ψ	$\|\overrightarrow{V} \cdot \overrightarrow{N_\perp}\|$	T
5000m	10°	30m/s	0.01s

According to Eq. (3), the visual resolution under current condition could be calculated as follows:

$$\begin{aligned}\Delta S &= 2l(x,y)L\cos^2\psi/d + \|\overrightarrow{V} \cdot \overrightarrow{N_\perp}\|T \\ &= \frac{2 \times \sqrt{\frac{5.76\times10^{-3}\times4.29\times10^{-3}}{7\times10^6}} \times 5000 \times \cos^2 10}{5.8 \times 10^{-3}} \\ &\quad +30 \times 0.01 \\ &= 3.44m.\end{aligned}$$

The simplified model controlled by the visual resolution is shown in Figure 5. Contrast to the original 3D model which contains large numbers of vertexes and triangle patches, the simplified model is more coincident with the human visual mechanism and easier for the calculation during succedent processing.



(a) original model (b) simplified model

Figure 5. The original model and the simplified model.

39 aspect graphs have been generated after the viewpoint space partition to the simplified model. Select the threshold $t = 0.17$, 17 aspect graphs will be gained after the clustering process.

7.2. 3D object recognition

Figure 6 illustrates the segmentation results of the object from farness to nearness in the simulation scene.

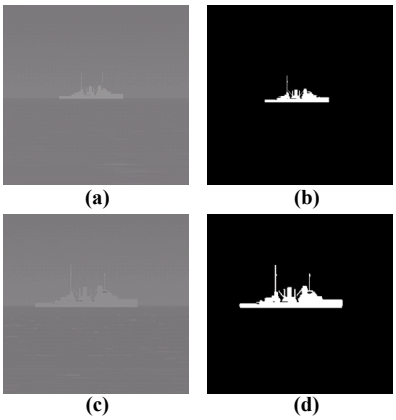


Figure 6. The simulation scenes and segmentation results. (a) is a remote simulation scene. (b) is a close simulation scene. (c), (d) are the segmentation results of (a), (b).

After the aspect graph library foundation, the segmentation object will be sent to match with the aspect graphs of different models. Take ship 1 for example, the minimum distances based on Fourier descriptor between ship 1 and five models are listed in Table 2. According to the minimum distance, the object of ship 1 is successfully recognised as the ship 1 model.

Table 2. The matching result

	plane 1	plane 2	ship 1	ship 2	car
ship 1	0.0038	0.0158	0.0012	0.0014	0.0032

Further more, select 30 discretional images of ship1, ship2 in the simulation scene and match them with the aspect graphs in the corresponding library of each model. The holistic recognition results of ship1, ship2 are 26/30 and 22/30, which signify the effect of the segmentation and recognition method proposed in this paper.

8. Conclusion

In this paper, existing segmentation and recognition methods in automatic navigation have been summarized and analyzed, and some new ideas have been proposed. An effective segmentation method based on nautical scene is first introduced in this paper. Simultaneously, a visual equation is employed to calculate the visual resolution, which would be used to control the 3D model simplification degree. A novel clustering method based on a tree structure is used to merge the similar aspect graphs generated by the simplified 3D model. Finally the segmented object will be sent to compare with the aspect graphs in the library according to a measurement of Fourier descriptor. The experiment results demonstrate that the segmentation and recognition method introduced in this paper could be applied in automatic navigation.

References

[1] P.J. Withagen, K.Schutte, A.M. Vossepoel, and M.G.Breuers. Automatic classification of ships from infrared (FLIR) images. *Proceedings of SPIE*, 3720:180–187, July 1999.

[2] J. Alves, J. Herman, and N.C. Rowe. Robust recognition of ship types from an infrared silhouette. *Command and Control Research and Technology Symposium, San Diego*, June 2004.

[3] V. Gouaillier and L. Gagnon. Ship silhouette recognition using principal components analysis. *Proceedings of SPIE*, 3164:59–69, October 1997.

[4] H. Hoppe, T. DeRose, and T. Duchamp. Mesh optimization. *Computer Graphics(Siggraph Proceedings 93)*, 19–26, January 1993.

[5] Y. Luo, H.M. Ma, and F.T. Li. An improved algorithm of viewpoint space partition in 3D object recognition application. *Proceedings of the 8th International Conference on Signal And Image Processing*, 34–39, August 2006.

[6] H.J.A.M. Heijmans. Composing morphological filters. *IEEE transactions on image processing*, 6(5):713–723, May 1997.

[7] J.A. Stark. Adaptive image contrast enhancement using generalizations ofhistogram equalization. *IEEE transactions on image processing*, 9(5):889–896, May 2000.

[8] N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[9] N. Aggarwal and W.C. Karl. Line detection in images through regularized hough transform. *IEEE transactions on image processing*, 15(3):582–591, March 2006.

[10] A.G. Bennett and R.B. Rabbetts. Clinical visual optics, third edition. *London:Butterworth Heinemann*, 20-22, 1998.

[11] D.S. Zhang and G.J. Lu. Review of shape representation andd escription techniques. *Pattern Recognition*, 37(1):1-19, 2004.

[12] M.J.de Ruiter and F.van Keulen. Topology optimization using a topology description function. *Structural and Multidisciplinary Optimization*, 26(6):406-416, April 2004.

[13] A. Folkers and H. Samet. Content-based image retrieval using Fourier descriptors on a logo database. *Proceedings of the 16th International Conference on Pattern Recognition*, 3:521-524, 2002.

[14] Princeton, http://shape.cs.princeton.edu/search.html, Princeton is a standard 3D model library, website accessed: 5 October 2007.

[15] M. Woo, J. Neider, T. Davis and D. Shreiner. OpenGL Programming Guide. *Addison-Wesley*, 1999.