2차원 모양 정보를 이용한 3차원 물체 검색 시스템

임 삼, 추 현 곤, 최 민 석, 김 회 율
한양대학교 전자통신전파공학과
전화 : 02-2290-0561 / 핸드폰 : 011-9769-1311

3D Object Retrieval System Using 2D Shape Information

Sam Lim, Hyon-Gon Choo, Min-Seok Choi, Whoi-Yul Kim
Division of Electrical and Computer Engineering, Hanyang University
E-mail : slim7538@vision.hanyang.ac.kr

Abstract

In this paper, we propose a new 3D object retrieval system using the shape information of 2D silhouette images. 2D images at different view points are derived from a 3D model and linked to the model. Shape feature of 2D image is extracted by a region-based descriptor. In the experiment, we compare the results of the proposed system with those of the system using curvature scale space(CSS) to show the efficiency of our system.

I. Introduction

With the development of the internet and network systems, increasing usage of digital images and very large volume of image databases naturally give rise to a difficult problem of organizing them and rapid access to their information contents. Thanks to the improvement of computing power, many retrieval systems have been developed to search 3D objects. Most methods use 3D information such as 3D spectrum and 3D surface normal vectors in order to obtain the features for 3D objects.[1] These methods, however, require high computational complexity and the constraint that user has to provide 3D data information to retrieve 3D objects. As an alternative, Karsten. etc[2] proposed 3D object retrieval method using 2D CSS Descriptor, but this method also suffers from the constraint of single contour requirement for an object.

In this paper, a new 3D object retrieval system which utilize the shape information of 2D silhouette images of a 3D object is proposed. Zernike moment(ZM) is used for computing the shape information of those 2D images.

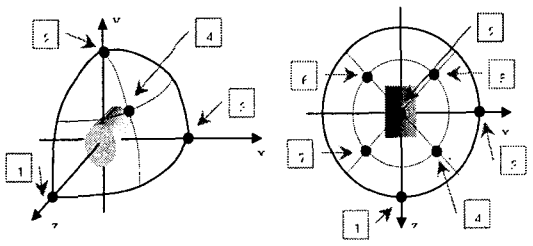
The rest of this paper is organized as follows. Section 2 shows how to construct 2D silhouette images of a 3D object and to extract the features of the object. In Section 3, the similarity measure method is described between the features of 3D query object and those of a object in the database. The experimental results are presented in section 4. and the conclusion is followed

II. Feature Extraction

Extracting features from a 3D object is as follows. First, a set of 2D silhouette image is constructed to represent a 3D object. From these 2D images, a set of shape features such as Zernike moments of silhouettes is extracted to represent a 3D object.

2.1 Construction of the 2D Silhouette Images of a 3D Object

It is important to determine view points of 2D image, which describe cleverly the property of a 3D object. Seven different views for a 3D object are extracted by analysing covariance matrix of the 3D object[3]. The eigen vectors of this matrix yield in the three orthogonal main directions. The views are named "Primary", "Secondary", and "Ternary" view according to the size of the eigenvalues. Another four views are added to cover the diagonal view directions in all quadrants of the hemisphere[3]. Fig. 1 shows the seven view directions of a 3D object.



(a) (b)

Fig. 1 (a) Three orthogonal main directions and a diagonal direction in the first octant, (b) four diagonal and main directions in the hemisphere y>0.

Along the different viewing directions, 2D silhouette images are generated by 2D projection. In Fig. 2, an example of 3D model on top and its associated seven 2D silhouette images are shown.

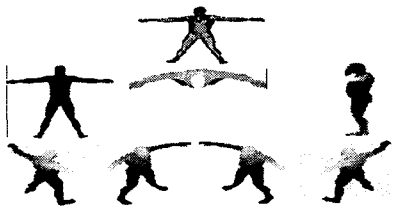


Fig 2. A 3D model and its seven silhouette images.

2.2. 2D Feature Extraction

From the generated silhouette images, 2D shape information is extracted by using Zernike moment[4]. Zernike moment has properties of rotation, scaling and translation invariance and multi-level representation[5]. The processes of extracting features are as follows.

step1: Silhouette images are generated.

step2: Since the Zernike moments are defined over a unit disk, the radius R of a circle is determined to enclose the shape completely from the centroid of the binarized shape in the image to the outer most pixel of the shape.

step3: The shape is then resampled to normalize to the size of 2R x 2R pixels. A typical value for R is 50. This normalization step allows the scale invariance for the descriptor.

step4: Zernike moment is extracted from the normalized image, and the magnitudes of Zernike moment are used as descriptor values.

III. Similarity Measurement

The similarity between two 3D objects are calculated by the 2D shape information of their silhouette images. The distance between 2D shape features of query image and those of images in the datebase are defined as in Eq. (1)

$$D = \sum_i \left\| M_d[i] - M_q[i] \right\| \quad (1)$$

$M_d[i]$:Magnitude of Zernike moment of images in DB

$M_q[i]$:Magnitude of Zernike moment of query image.

Each view of the query object(A) is compared to those of other objects(B) in database[3]. The overall matching error is computed by summing the errors from all possible views. This procedure can be described by the error matrix between objects A and B, where the matrix element $e_{k,l}$ represents the similar distance between kth view of the object A and 1st view of the object B. The following example is the case when the number of multi-views for object A and B are three.

$$\begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} \\ e_{2,1} & e_{2,2} & e_{2,3} \\ e_{3,1} & e_{3,2} & e_{3,3} \end{bmatrix}$$

All possible combinations are as follows.

$e_{1,1} + e_{2,2} + e_{3,3}$, $e_{1,1} + e_{2,3} + e_{3,2}$, $e_{1,2} + e_{2,1} + e_{3,3}$, $e_{1,2} + e_{2,3} + e_{3,1}$, $e_{1,3} + e_{2,1} + e_{3,2}$, $e_{1,3} + e_{2,2} + e_{3,1}$

In this case, the minimum value is selected as the similarity measure between object A and B.

Ⅳ. Experiment

For the experiment, we used 227 3D models which are expressed in VRML and its 1589 silhouette images, which are used for MPEG-7 core experiment procedure. Fig. 3 shows the examples of 3D images in the dataset.

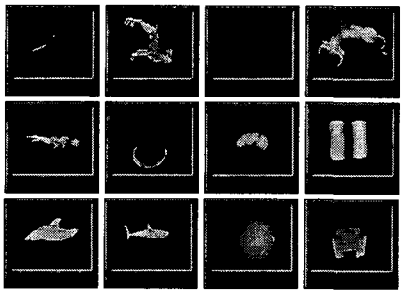


Fig. 3 Examples of 3D model

We classified 227 3D models and their silhouette images into 15 groups. Table 1 shows the types and the numbers of classified images. BEP(Bull's Eye Performance)[6] is used as retrieval rate measurement. For the evaluation of the performance of the proposed system, we compared with the system employed CSS proposed in MPEG-7[2].

Table 1. Database list of 3D cafe

Class Name	Number	Class Name	Number
Aerodynamic	35	Letter_b	10
Ballon	7	Letter_c	10
Building	10	Letter_d	10
Car	17	Letter_e	10
Elm	9	Missile	10
Finger	30	Soma	7
Fourlimb	31	Tree	21
Letter_a	10		

Fig. 4 shows the results of retrieval efficiency for the number of 2D multi-views. The more the number of multi-views is, the higher the retrieval rate. In most cases, however, the retrieval rates with the ZM descriptor is much better than that with CSS descriptor.

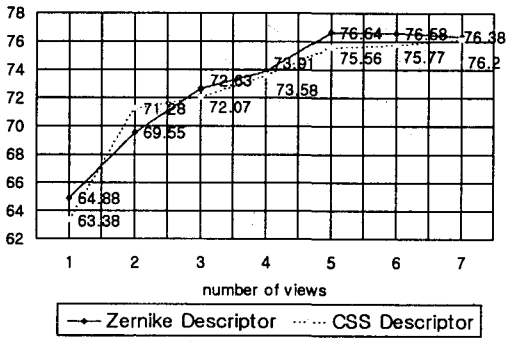


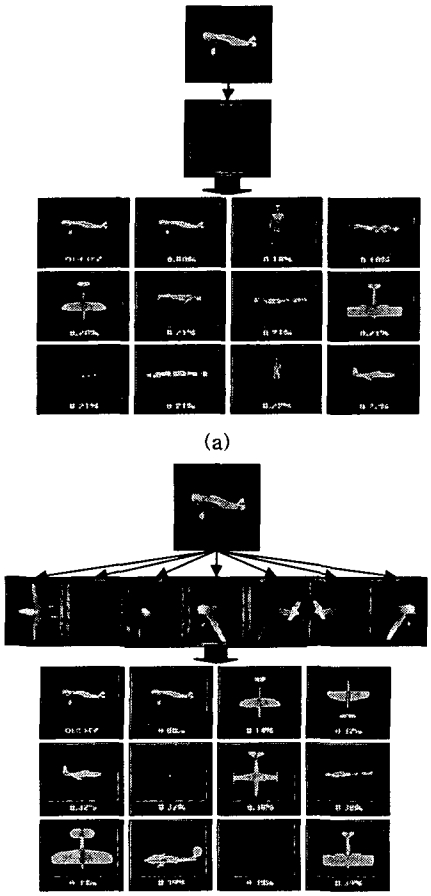
Fig. 4 Results of retrieval rate v.s. the number of views

When the restricted views are used, Table 3 shows the result for the selection of viewing directions. The numbers denote view directions. Case 1 shows the retrieval rate when three main views are selected, and case 2 shows the retrieval rate when the rest views except the main directions are used. The selection of the main views have little influence over the overall retrieval rate. When the rest of views are used except the main views, the retrieval rate decrease considerably.

Table 3. Results depending on the selection of different view points.

Case 1	Average BEP(%)	Case 2	Average BEP(%)
1, 2	69.55	4, 5, 6	65.18
1, 3	68.16	4, 5, 7	67.20
2, 3	68.17	4, 6, 7	66.57
1, 2, 3	72.63	5, 6, 7	65.03

In Fig. 5, we visually show the result that the retrieval rate with seven views is much higher than that with a single view.



(a)

(b)

Fig. 5. 3D objects retrieval results
(a) using a single view
(b) using multiple views

V. Conclusion

In this experiment, the retrieval rate using ZM descriptor was far better than that of using CSS descriptor. However, the following observation can be made for both ZM and CSS descriptor. The more the number of multiple views was used, the higher the retrieval rate. But we noticed that the complexity increased exponentially as the number of multiple views increased. Therefore both issues of the complexity and the retrieval rate must be considered in designing the 2D-3D retrieval system.

Reference

[1] MPEG-7 Visual part of experimentation model Version 10.0, ISO/IEC JTC1/SC29/WG11/N4063, pp.51-55, Singapore, March, 2001.

[2] Karsten Muller, (HHI), Jens-Rainer Ohm(HHI), James Cooper (Mitsubishi), Miroslaw Bober (Mitsubishi), "Results of 2D/3D Shape Core Experiments MS-4 ", ISO/IEC JTC1/SC29/ WG11/MPEG/M6190 July 2000, Beijing.

[3] MPEG-7 Visual part of experimentation model Version 10.0, ISO/IEC JTC1/SC29/WG11/N4063, pp.11-15, Singapore, March, 2001.

[4] Yong-Sung Kim, "Shape Descriptors for Content-Based Image Retrieval", Ph. D. thesis, pp. 41-45, Hanyang University, 2000.

[5] Whoi-Yul Kim and Yong-Sung Kim, "A Rotatio Invariant Geometric Shape Description Using Zernike Moments", ISO/IEC JTC1/SC29/WG11, P687, Lancaster, February 1999.

[6] S. Jeannin, M. Bober, "Description of core experiments for MPEG-7 motion/shape", ISO/IEC JTC1/SC29/WG11/ MPEG99 /N2690, 1999.