Short Paper

Shape Memorization and Recognition of 3D Objects Using a Similarity-Based Aspect-Graph Approach*

JWU-SHENG HU AND TZUNG-MIN SU
Department of Electrical and Control Engineering
National Chiao Tung University
Hsinchu, 300 Taiwan

This work proposes an incremental combinational algorithm to generate the prototype of a 3D object using 2D images randomly sampled from a viewing sphere. Similarity-based aspect-graph, which contains a set of aspects and prototypes for these aspects, is employed to represent the database of 3D objects. Furthermore, the proposed algorithm is based on low-level features and similarity measures between the features. In this work, the Fourier descriptor and point-to-point lengths are adopted as features, and three similarity measures, called the 1-norm, 2-norm, and K-L distance, are adopted to extract characteristic views. The effectiveness of the proposed algorithm is demonstrated by experiments with an updating mechanism.

Keywords: aspect-graph, fourier descriptor, object recognition, shape memorization, similarity measure

1. INTRODUCTION

Object recognition is an important topic in computer vision owing to its applicability in many applications, including mobile robot localization and navigation, visual servoing, surveillance and military applications. Although various approaches have been developed to solve the recognition problem [1, 2], object recognition remains a difficult task in the field of computer vision, especially for 3D object recognition. Variations in viewing direction and angle, illumination changes, scene clutter and occlusion make 3D object recognition difficult and impractical in real-world applications.

3D object recognition differs from the variations of the position and type of the illumination source, or the relative position of observer and object. Therefore, some high-level theorems of 3D object perception have been studied to solve the above-mentioned weakness, improving the practicality of the 3D object recognition task [3]. Existing theorems about the high-level 3D object perception can be categorized as object-center and viewer-center representations based on the coordinate system [1], and as volume-based (or model-based) and view-based representation based on the constituent elements [4]. A viewer-centered representation describes the parts of an object relative to a coordinate

Received April 20, 2007; revised August 23, 2007 & March 3, 2008; accepted April 10, 2008.
Communicated by Jenq-Neng Hwang.

* This paper was presented on IEEE International Conference on Systems, Man and Cybernetics, October 8-11th, 2006, Taipei, Taiwan.

system based on the observer. A view-based representation represents a 3D object with a set of object views.

The traditional aspect-graph method [5, 6] assumes that an object belongs to a limited class of shapes, and that characteristic views can be extracted using prior knowledge of the object. Cyr and Kimia [7] presented a similarity-based aspect-graph method, using curve matching and shock matching as the similarity measures between two views, to improve the practicability of aspect-graph representation for extracting the characteristic views. However, the process of updating the aspect-graph representation is inconvenient, since the training views of an object in [7] are sampled at 5-degree increments. To improve the flexibility of the update mechanism, a similarity-based aspect-graph representation that can be constructed with object views sampled at random intervals is studied in this work. Furthermore, the object representation becomes increasingly accurate after gathering more new object views with limited increasing search space using the proposed combinational algorithm. Additionally, this study utilizes three different similarity measures from the similarity measurements used in [7], called 1-norm, 2-norm and K-L distance, to reduce the computing time. Furthermore, to represent rigid objects efficiently, Fourier descriptors of the sampled points on the object contour and point-to-point lengths are calculated based on the shape contour in this work. Fig. 1 illustrates the block diagram of the overall scheme, where T_1 denotes the minimum number of sampled 2D views adopted to extract the object prototype.

The remainder of this paper is organized as follows. Section 2 presents the procedure of extracting the main (Fourier descriptors) and assistant (point-to-point lengths) features, which are utilized to measure the similarity between two views. Section 3 describes the novelty of this work, namely the construction and updating of similarity-based aspect-graph representations from a set of objects views sampled at random intervals, and the procedure of recognizing 3D objects from 2D object views. Section 4 presents experimental results to demonstrate the performance of the proposed method. Conclusions are finally drawn in section 5.

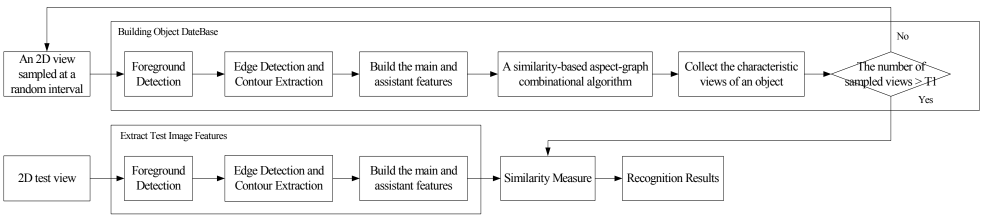

Fig. 1. Basic workflow of the proposed framework; T_1 denotes the number of sampled views required to build the aspect-graph representation of an object.

2. FEATURE EXTRACTION

2.1 Foreground Detection and Contour Extraction

Shadows and highlights should be removed before extracting the object features to eliminate the effects of lighting. A robust background subtraction framework from our previous work [8, 9] is adopted to extract the foreground regions with the consideration of shadows and highlights. This study utilizes the shape feature to measure the similarity between two object views. To extract the shape information from the foreground object, Canny edge detection [10] is adopted to obtain the shape edge, and Gradient Vector Flow Snake [11] is then applied to extract the contour information. The contour information is included in a set Z, which consists of N points z_i, where z_i can be described as a complex form as in Eq. (1).

$$Z = \{z_i\} = \{x_i + jy_i\},\ 0 \le i < N \tag{1}$$

2.2 Generation of the Main and Assistant Feature

To avoid variations in shift and scale, the points inside the set Z are re-sampling using Eq. (2). The Fourier transform is applied on $\tilde{Z}$ to derive the Fourier descriptors. The first T_2-magnitude parts are extracted as the main feature to describe the object shape without the variations on the high-frequency noises. The method for extracting the main feature F_m is described as in Eq. (3).

$$\tilde{Z} = \{\tilde{z}_i\} = \{(\tilde{x}_i + j\tilde{y}_i\} = \{L_c[(x_i - x_c) + j(y_i - y_c)]/L\} \tag{2}$$

where $0 \le i < N, L$ denotes the real contour length, and L_c denotes the expected contour length.

$$F_m = \{|f_t|, |f_{N-t}|, 0 \le t \le T_2\} \tag{3}$$

where $|f_t|$ and $|f_{N-t}|$ denotes the magnitude part of Fourier descriptor at frequency t.

Moreover, to consider take the details of an object, the lengths between each two points of the set $\tilde{Z}$ are calculated as the assistant feature F_a, which is described as Eq. (4).

$$F_a = \{l_i\} = \{\| \tilde{z}_i - \tilde{z}_{i-1} \|\} = \{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}\} \tag{4}$$

where $0 \le i < N$.

3. FEATURE EXTRACTION

3.1 Similarity Function

To determine the similarity between two contours, a similarity measurement has to be applied on the extracted two contours, which each consist of N points. Assume that the features extracted from two contours are respectively $U = \{u_0, \ldots, u_i, \ldots, u_{N-1}\}$ and $V = \{v_0, \ldots, v_i, \ldots, v_{N-1}\}$. The three similarity measures are then calculated using Eqs. (5)-(7).

- 1-norm distance: $D_{1-norm}(u, v) = \sum_{i=0}^{N-1} |u_i - v_i|$ (5)
- 2-norm distance: $D_{2-norm}(u, v) = \sqrt{\sum_{i=0}^{N-1} (u_i - v_i)^2}$ (6)

• K-L distance [12]:

$$D_{KL}(p_1 \parallel p_0) \approx \sum_{t=0}^{N-1}\left(p_1(t)\cdot\log(\frac{p_1(t)}{m(t)}) + p_0(t)\cdot\log(\frac{p_0(t)}{m(t)})\right)$$
$$p_0 = \frac{U}{u_{sum}},\ p_1 = \frac{V}{v_{sum}},\ u_{sum} = \sum_{i=0}^{N-1} u_i,\ v_{sum} = \sum_{i=0}^{N-1} v_i \tag{7}$$
$$m(t) = \frac{p_0(t)+p_1(t)}{2}$$

3.2 Generation of Aspects and Characteristic Views

Cyr and Kimia presented a combinational algorithm [7] (includes two criterions, called local monotonicity and object-specific distinctiveness of aspect views) to generate the aspects and characteristic views via collected 2D views sampled at 5-degree increments in order. If more views of an object are captured to improve the object representation in the work of [7], the total views of an object have to be resorted in order of captured angles to extract the new characteristic views and aspects. The first criterion in the work of [7], called the local monotonicity, might not be matched when the object itself is symmetric in some level or the object becomes symmetric in the feature space.

This work proposes an incremental combinational algorithm motivated by the work of [7] to solve the above problems. The aspects of 3D objects can be extracted using 2D images sampled at random intervals. The object representation becomes increasingly detailed using new captured views. Besides, characteristic views are extracted without re-calculating the similarity measures by re-sorting total collected views. Such an incremental combinational algorithm provides a supervised learning ability for related applications in an uncontrolled environment. For example, a home robot memorizes a 3D object well when it gets more and more 2D views of the 3D object at different times. Furthermore, the proposed algorithm extracts characteristic views by measuring the similarity between a new 2D view and the extracted characteristic views. The relationship between the sorted neighbor views is not considered in the proposed algorithm. Thus, a local minimum region caused from the first criterion in the work of [7] is not necessary for extracting the aspects. Although the proposed approach cannot verify the test view with the specific view of an object, it improves the flexibility of building aspect-graph representation, and reduces the computing time of updating object aspects.

Assume that V_{new}^n denotes the new sampled view of the n^{th} object; C_m^n denotes the m^{th} characteristic view of the n^{th} object; $C_{m^{\min}-1}^n$ and $C_{m^{\min}+1}^n$ denote the neighboring views with a minimum distance between V_{new}^n, and $m^{\min}$ denotes the index of the aspect that has the minimum distance with V_{new}^n. Aspects and characteristic views are then imposed using three steps, shown as steps A-1 to A-3. Fig. 2 shows the flowchart of the proposed aspect-graph representation.

Step A-1: When the number of existing aspects equals zero, V_{new}^n is considered as a characteristic view of a new aspect.

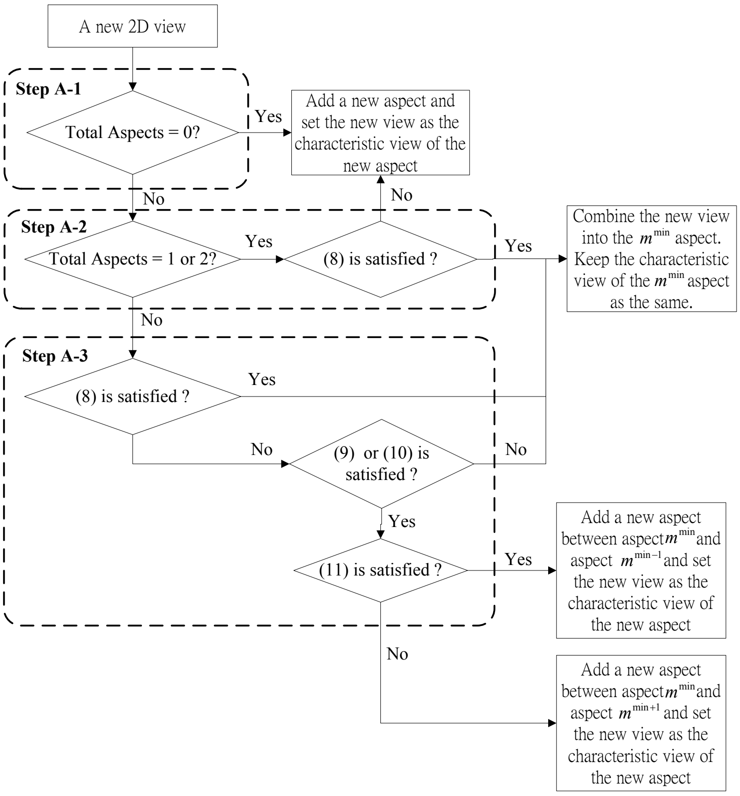

Fig. 2. The procedure of the proposed combinational algorithm procedure.

Step A-2: When the number of existing aspects equals 1 or 2, V_{new}^{n} is combined into one existing aspect and the characteristic views of the aspect remains the same if Eq. (8) is satisfied. Otherwise, a new aspect is generated, and V_{new}^{n} is considered as the characteristic view of the new aspect.

$$\min_{\text{all } C_m} d(V_{new}, C_m) < T_3 \tag{8}$$

where T_3 denotes a predefined threshold value.

Step A-3: When the number of existing aspects is greater than or equal to 3, a new aspect is built, and V_{new}^{n} is regarded as the characteristic view of the new aspect if Eqs. (9) or (10) are satisfied. Otherwise, V_{new}^{n} is combined into aspect $m^{\min}$, which retains its previous characteristic view.

$$\min_{\text{all } C_m} d(V_{new}, C_m) > T_4 \tag{9}$$

$$T_3 \le \min_{\text{all } C_m} d(V_{new}, C_m) \le T_4 \text{ and } d(V_{new}, C_{m^{\min} \pm 1}) > T_4 \tag{10}$$

where T_4 denotes a predefined threshold value.

Furthermore, if a new aspect is built, then the aspect order can be determined using Eq. (11). If the similarity distance between V_{new}^{n} and $C_{m^{\min}+1}^{n}$ is greater than the similarity distance between V_{new}^{n} and $C_{m^{\min}-1}^{n}$, then the new aspect is inserted between aspect $m^{\min}$ and aspect $m^{\min-1}$. Otherwise, the new aspect is inserted between aspect $m^{\min}$ and aspect $m^{\min+1}$. Therefore, the similar aspects are close to each other.

$$d(V_{new}, C_{m^{\min}+1}) > d(V_{new}, C_{m^{\min}-1}) \tag{11}$$

Terms T_3 and T_4 denote two predefined threshold values, where $T_4 > T_3$. The criterion for selecting T_3 and T_4 depends on the feature selection, and on the precise level for describing the object. If T_3 and T_4 are both small, then the criterion of combining 2D views becomes strict and thus the number of aspect becomes more numerous. Furthermore, if the difference between T_3 and T_4 decreases, then the tolerance of difference between 2D views inside an aspect decreases, thus increasing the number of aspects. Furthermore, T_3 and T_4 should be initialized manually and modified iteratively until the final number of aspect reaches an acceptable level, determined by the symmetry of the object. In this work, T_{31} and T_{41} are defined as the T_3 and T_4 of the main feature (Fourier descriptors), and T_{32} and T_{42} are defined as the T_3 and T_4 of the assistant feature (point-to-point lengths). Section 4 presents the values of T_{31}, T_{41}, T_{32}, and T_{42}.

3.3 Object Recognition Using 2D Characteristic Views

After constructing the aspect-graph representation of each object in the database (Fig. 2), a test view of an unknown object can be recognized using the similarity measure with the main features and the assistant features.

Step B-1: The test 2D view of an unknown object is compared with 2D characteristic views using the main features. The first T_5 2D characteristic views in the database having the smallest similarity distance with the test 2D view are then preserved for further recognition.

Step B-2: Suppose that T_6 objects are included in the selected T_5 characteristic views, the final similarity distances can be calculated from the assistant features of these T_6 objects using Eq. (12).

$$d(V_j^i, V_m^n) = d_{\text{assistant}}(V_j^i, V_m^n) + (T_4 / T_3) \times d_{main}^{\min}(n) \tag{12}$$

where V_j^i denotes the views of an unknown object; V_m^n denotes view m of object n, and $d_{\text{main}}^{\min}(n)$ denotes the minimum similarity distance calculated using the main feature between the unknown object and the n^{th} object of the database.

4. EXPERIMENTAL RESULTS

This section describes several experiments that demonstrate the effectiveness of the proposed method. A SONY EVI-D30 PTZ camera was utilized to capture object views.

Fig. 3 displays the image database contains 12 real rigid objects. The training views of each object were captured with five-degree increment intervals, and were collected as $V_d^n = \{V^n(d)\}$, $1 \le n \le 12$, $1 \le d \le T_1$, with T_1 views for each object. The additional views of each object were captured from the trisection-points between each five-degree point, and were collected as $V_t^n = \{V^n(t)\}$, $1 \le n \le 12$, $1 \le t \le 216$, with 216 views for each object. Furthermore, in the following experiments, $T_1 = 72$, $T_2 = 25$, $T_{31} = 640$, $T_{32} = 800$, $T_{41} = 336$ and $T_{42} = 480$. Furthermore, T_5 is defined as the half of total characteristic views in the aspect-graph. The computing time taken to calculate the similarity between a test view and a view in the database was about 0.005 seconds with P4 3.2G CPU and 1GB RAM.

Fig. 3. Image database comprising twelve 3D rigid objects, where object 1, object 2, …, object 12 are listed from left to right and from top to bottom.

4.1 The Efficiency of Three Similarity Measures on Rigid Object Recognition

The first experiment shows the aspect numbers extracted using the algorithm proposed by Cyr and Kimia [7] and the recognition results using the extracted aspects and three similarity measures, 1-norm distance, 2-norm distance, and K-L distance. The views in V_t^n were then tested with the characteristic views of each aspects using three similarity measures. Table 1 shows the results of the aspect numbers of each object. Table 2 shows the recognition results. The results in Tables 1 and 2 indicate that the aspect-graph representation based on 1-norm distance generates the fewest aspects, and has the best recognition performance. Furthermore, the computing time of 1-norm distance was also the lowest among the three similarity measures. Therefore, for brevity, 1-norm distance was adopted to demonstrate the performance in the following experiments.

4.2 3D Rigid Object Recognition with Building Database Using 72 Object Views

In the second experiment, the efficiency of the proposed algorithm was measured using 2D views captured at random intervals. To determine the average performance of the proposed method, aspects were generated by sampling the views in 200 different random orders. For performing the efficiency of the proposed method, a comparison between

Table 1. Numbers of aspect using the main feature and aspect-graph representation in [7] with three similarity measures.

Numbers of aspect	The index of the objects in the database listed in Fig. 3											
	1	2	3	4	5	6	7	8	9	10	11	12
KL distance	48	29	38	34	29	32	38	29	42	35	32	32
1-norm distance	39	35	33	42	29	32	38	23	31	37	32	40
2-norm distance	38	34	51	50	33	39	37	22	27	47	39	49

Table 2. Result of matching unknowns using 1-norm, 2-norm and K-L distance with aspect-graph representation in [7].

Recognition Rate	In Top 1 Matches (%)	In Top 2 Matches (%)	In Top 3 Matches (%)
KL distance	97.61	98.30	98.61
1-norm distance	98.69	99.04	99.15
2-norm distance	97.34	97.69	97.80

Table 3. Numbers of aspect using the proposed method (M2) and that proposed by [7] (M1) with 1-norm distance.

Index of objects	Numbers of aspect with the main feature		Numbers of aspect with the assistant feature	
	M1	M2	M1	M2
object 1	39	34.7	37	38.7
object 2	35	3.8	40	14.1
object 3	33	27.8	34	14.3
object 4	42	24.8	29	22.8
object 5	29	6.9	35	11
object 6	32	9.5	41	20.1
object 7	38	2	39	8.4
object 8	23	25.6	34	31.1
object 9	31	17.1	26	25.8
object 10	37	16.2	32	17.7
object 11	32	16.6	36	23.6
object 12	40	28.8	26	19.9

the proposed combinational algorithm and that proposed by [7] is performed with the same dataset (rigid object, Fig. 3) with 1-norm distance. Table 3 shows the number of aspects, where M1 denotes the results performed with the algorithm proposed by [7] and M2 denotes the results of the proposed method. In Table 3, the number of aspects of symmetric object in M2 is fewer than those in M1, especially for objects 2, 5, 6 and 7. Moreover, Table 4 shows the recognition results. The views in V_t^n were adopted as unknowns, and tested whenever the aspect-graph representations were built each time (total 200 times). From Table 4, the average recognition results of the proposed method and that proposed by [7] are almost the same. Table 5 shows the statistical information about the mean and the standard deviation of the aspect number. The small standard deviation in Table 5 demonstrates that the proposed combinational algorithm generated the aspects without the effects of the training views order. Furthermore, Table 6 shows the statistical

Table 4. Result of matching unknowns using the proposed method (M2) and that proposed by [7] (M1) with 1-norm distance via main and assistant features.

Index of objects	Top 1 Matches (%)		Top 2 Matches (%)		Top 3 Matches (%)	
	M1	M2	M1	M2	M1	M2
object 1	95.83	98.25	97.22	99.21	97.69	99.61
object 2	100	99.97	100	100	100	100
object 3	99.07	97.71	99.07	98.96	99.07	99.39
object 4	97.22	97.39	98.61	98.73	99.07	99.34
object 5	100	100	100	100	100	100
object 6	100	99.81	100	99.96	100	99.98
object 7	100	99.78	100	99.87	100	99.89
object 8	98.15	99.35	98.61	99.67	98.61	99.78
object 9	100	99.90	100	99.97	100	99.99
object 10	98.61	97.97	99.54	98.68	99.54	98.98
object 11	100	98.44	100	99.47	100	99.77
object 12	95.37	96.83	95.37	98.17	95.83	98.65
Average	98.69	98.78	99.04	99.39	99.15	99.62

Table 5. Number of aspects using main and assistant features and the proposed aspect-graph generation with 1-norm similarity measure on rigid object recognition, where "Std." denotes standard deviation.

Numbers of aspect		The index of the objects in the database listed in Fig. 3											
		1	2	3	4	5	6	7	8	9	10	11	12
main feature	Mean	34.7	3.84	27.8	24.8	6.9	9.5	2.0	25.6	17.1	16.2	16.6	28.8
	Std.	1.82	0.56	1.62	1.07	0.82	0.96	0.49	1.28	1.79	1.24	1.00	1.04
assistant feature	Mean	38.7	14.1	14.3	22.8	11.0	20.1	8.4	31.1	25.8	17.7	23.6	19.9
	Std.	2.51	1.67	1.81	1.79	1.17	1.94	1.09	2.07	2.18	1.68	1.81	1.59

Table 6. Recognition rates using main and assistant features and the proposed aspect-graph generation with 1-norm similarity measure on rigid object recognition.

Recognition Rates		The index of the objects in the database listed in Fig. 3												
		1	2	3	4	5	6	7	8	9	10	11	12	Avg.
Top 1 (%)	Mean	98.3	100	97.7	97.4	100	99.8	99.8	99.4	99.9	98.0	98.4	96.8	98.8
	Std.	0.96	0.18	1.35	0.98	0.00	0.00	0.24	0.64	0.23	0.92	0.83	0.70	0.59
Top 2 (%)	Mean	99.2	100	99.0	98.7	100	100	99.9	99.7	100	98.7	99.5	98.2	99.4
	Std.	0.64	0.00	0.90	0.67	0.00	0.13	0.21	0.42	0.11	0.70	0.49	0.65	0.41
Top 3 (%)	Mean	99.6	100	99.4	99.3	100	100	99.9	99.8	100	99.0	99.8	98.6	99.6
	Std.	0.44	0.00	0.62	0.49	0.00	0.09	0.20	0.33	0.07	0.59	0.31	0.50	0.30

information about the recognition results, including the mean and the standard deviation of the recognition rate. From Table 6, the proposed aspect-graph generation is efficient owing to its high recognition rate and low standard deviation.

Table 7. The result of the numbers of aspect using the main and assistant features and the proposed aspect-graph representation in with 1-norm similarity measure.

Numbers of aspect	The index of object in the database											
	1	2	3	4	5	6	7	8	9	10	11	12
D_{18}	14.11	3.40	11.98	10.13	5.32	6.36	1.58	12.64	8.80	8.43	8.73	11.10
D_{36}	22.86	3.60	18.83	16.15	6.23	8.01	1.80	19.16	12.53	12.17	12.48	18.29
D_{54}	29.52	3.74	23.95	20.96	6.645	8.94	1.92	23.24	15.20	14.53	15.06	24.05
D_{72}	34.66	3.84	27.83	24.75	6.87	9.47	2.04	25.62	17.14	16.16	16.62	28.75
D_{90}	39.28	3.99	28.68	25.99	7.14	9.83	2.14	27.32	18.04	17.37	17.86	30.99
D_{108}	43.28	4.07	29.50	27.14	7.36	10.12	2.26	28.67	18.90	18.50	19.06	33.12

4.3 3D Object Recognition with Updating Database Using 18, 36, 54, 72, 90, 108 Views

The proposed method can construct the aspect-graph representation with sampled views at random intervals, making an updating mechanism that integrates the database using new collected views practicable. In this experiment, 18 random views sampled from V_d^n were first utilized to construct the coarse aspect-graph representation of each object, called D_{18}. Eighteen additional random views were then taken from the remaining views in V_d^n to update the coarse database D_{18} to a more accurate database, called D_{36}. Similarly, D_{54} and D_{72} were constructed using views in the remainders of V_d^n . Moreover, D_{90} and D_{108} were further constructed with extra random views sampled from V_t^n. Table 7 shows the average aspect numbers of each rigid object from 200 runs. Although the aspect numbers increased when new views were employed to update the coarse database, the number of stored views was still much smaller than the number of original views. Fig. 4 shows the results of recognition rate using coarse to fine databases, and Fig. 5 shows the results of the standard deviation of recognition rate. As shown in Fig. 4, the recognition rate rises when the aspect-graph representations are trained from more object views. Moreover, the stability becomes better according to the decreasing standard deviation. Therefore, the proposed method is demonstrated to be useful for updating the aspect-graph representations without re-sorting the overall collected views, or re-calculating the overall similarity measures.

5. CONCLUSIONS

This study proposes a flexible combinational algorithm to build the aspect-graph representation with 2D images sampled at random intervals for recognizing 3D objects. Fourier descriptor and the length between each point on the object contour are adopted as the main and assistant features, respectively, to measure the similarity between each object view using three similarity measures, namely the 1-norm, 2-norm and K-L distances. Although the relationships between the aspects are lost, the number of aspects for the symmetrical objects is reduced and flexibility of the database update mechanism increases without the need to reorder the overall collected views. Therefore, the proposed algorithm improves the computing time while updating the database. A new view of an object can be adopted to increase the integrity of the database using the proposed algorithm. Furthermore, to make the databases easy to construct, foreground detection with shadow and highlight removal is adopted to extract the interested object. Although the combination of proposed main feature and assistant feature is valuable for representing objects, the proposed method has a high computation requirement, and improving its efficiency is a topic for future work.

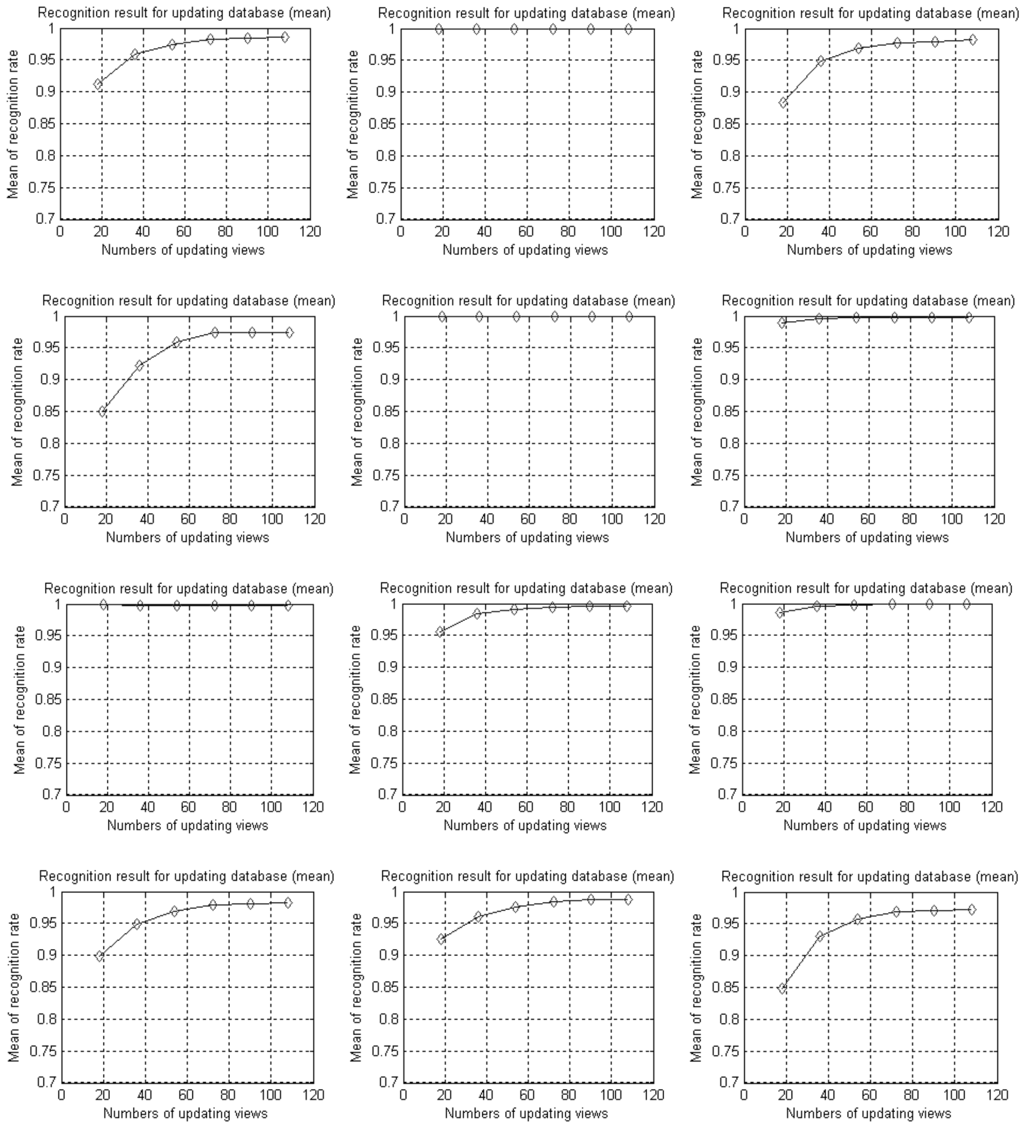

Fig. 4. Recognition rates of coarse and fine databases (D_{18}, D_{36}, D_{54}, D_{72}, D_{90} and D_{108}), calculated with 200 results.

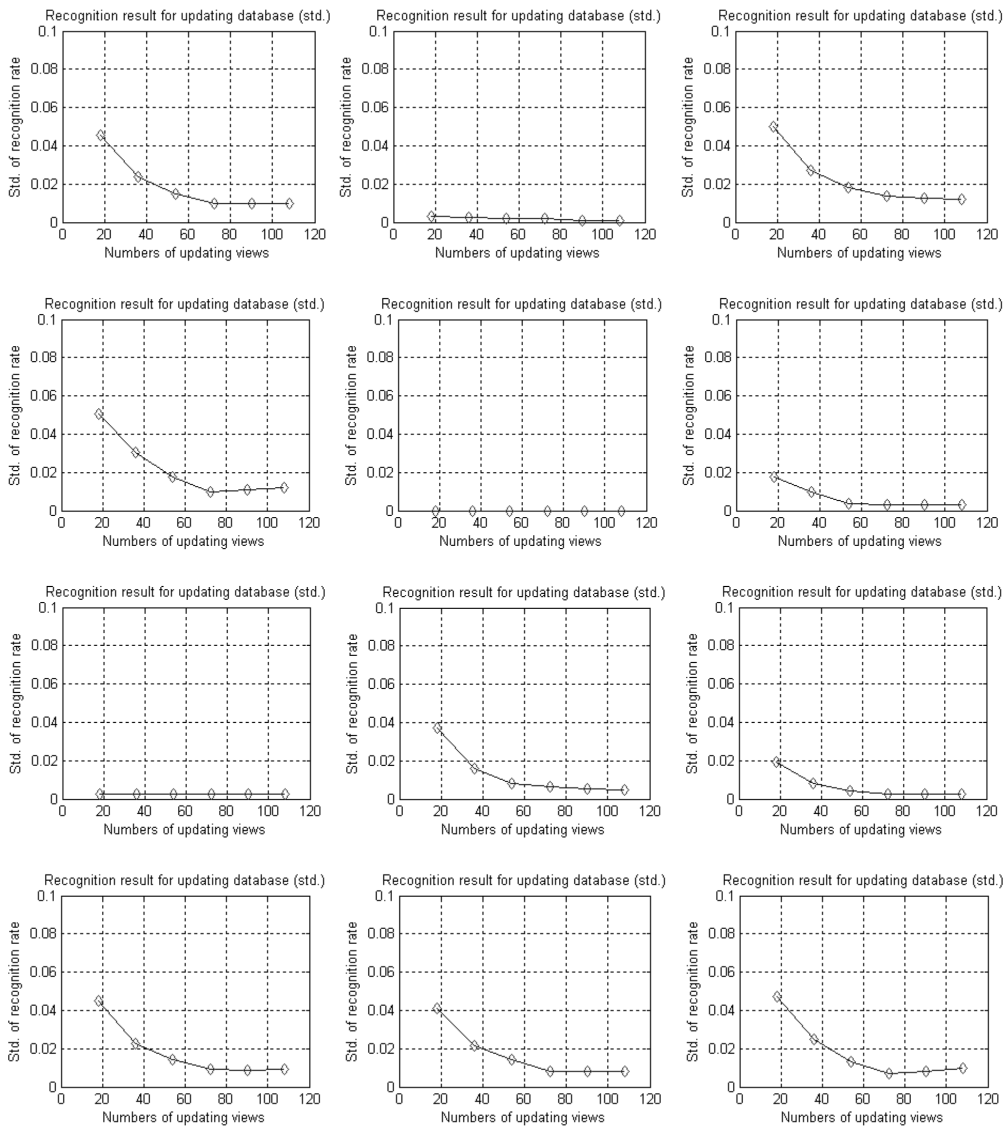

Fig. 5. Standard deviations of recognition rates using coarse to fine databases (D_{18}, D_{36}, D_{54}, D_{72}, D_{90} and D_{108}), calculated with 200 results.

REFERENCES

1. G. Peters, "Theories of three-dimensional object perception-a survey," *Recent Research Developments in Pattern Recognition*, Transworld Research Network, Kerala, India, Vol. 1, 2000, pp. 179-197.
2. G. Mamic and M. Bennamoun, "Representation and recognition of 3D free-form objects," *Digital Signal Processing*, Vol. 12, 2002, pp. 47-76.

3. V. Blanz, M. J. Tarr, and H. H. Büulthoff, "What object attributes determine canonical views?" *Perception*, Vol. 28, 1999, pp. 575-600.
4. I. Weiss and M. Ray, "Model-based recognition of 3D objects from single images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 116-128.
5. J. J. Koenderink and A. J. van Doorn, "The singularities of the visual mapping," *Biological Cybernetics*, Vol. 24, 1976, pp. 51-59.
6. I. Shimshoni and J. Ponce, "Finite-resolution aspect graphs of polyhedral objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 315-327.
7. C. M. Cyr and B. B. Kimia, "A similarity-based aspect-graph approach to 3D object recognition," *International Journal of Computer Vision*, Vol. 57, 2004, pp. 5-22.
8. J. S. Hu, T. M. Su, and S. C. Jen, "Robust background subtraction with shadow removal for indoor environment surveillance," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 4545-4550.
9. J. S. Hu, T. M. Su, and C. C. Lin, "Shape memorization and recognition of 3D objects using a similarity-based aspect-graph approach," in *Proceedings of the IEEE International Conference on Systems*, *Man*, *and Cybernetics*, 2006, pp. 4920-4925.
10. J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, 1986, pp. 679-698.
11. C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 66-71.
12. Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, Vol. 40, 2000, pp. 99-121.

Jwu-Sheng Hu (胡竹生) received the B.S. degree from the Department of Mechanical Engineering, National Taiwan University, Taiwan, in 1984, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering, University of California at Berkeley, in 1988 and 1990, respectively. He is currently a Professor in the Department of Electrical and Control Engineering, National Chiao Tung University, Taiwan, R.O.C. His current research interests include microphone array signal processing, active noise control, intelligent mobile robots, embedded systems and applications.

Tzung-Min Su (蘇宗敏) was born in 1978. He received the B.S. degree in Electrical and Control Engineering from National Chiao Tung University, Taiwan, R.O.C. in 2000. He is currently a Ph.D. candidate in Department of Electrical and Control Engineering at National Chiao Tung University, Taiwan, R.O.C. He is the championship of the national competition held by Ministry of Education Advisor Office in 2001. His research interests include background subtraction, 3D object recognition, and home-care surveillance.