Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 89691, 9 pages
doi:10.1155/2007/89691

Research Article

Content-Based Object Movie Retrieval and Relevance Feedbacks

Cheng-Chieh Chiang,[1,2] Li-Wei Chan,[3] Yi-Ping Hung,[4] and Greg C. Lee[5]

[1] *Graduate Institute of Information and Computer Education, College of Education, National Taiwan Normal University, Taipei 106, Taiwan*
[2] *Department of Information Technology, Takming College, Taipei 114, Taiwan*
[3] *Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Taiwan University, Taipei 106, Taiwan*
[4] *Graduate Institute of Networking and Multimedia, College of Electrical Engineering and Computer Science, National Taiwan University, Taipei 106, Taiwan*
[5] *Department of Computer Science and Information Engineering, College of Science, National Taiwan Normal University, Taipei 106, Taiwan*

Received 26 January 2006; Revised 19 November 2006; Accepted 13 May 2007

Recommended by Tsuhan Chen

Object movie refers to a set of images captured from different perspectives around a 3D object. Object movie provides a good representation of a physical object because it can provide 3D interactive viewing effect, but does not require 3D model reconstruction. In this paper, we propose an efficient approach for content-based object movie retrieval. In order to retrieve the desired object movie from the database, we first map an object movie into the sampling of a manifold in the feature space. Two different layers of feature descriptors, dense and condensed, are designed to sample the manifold for representing object movies. Based on these descriptors, we define the dissimilarity measure between the query and the target in the object movie database. The query we considered can be either an entire object movie or simply a subset of views. We further design a relevance feedback approach to improving retrieved results. Finally, some experimental results are presented to show the efficacy of our approach.

1. INTRODUCTION

Recently, it has been more popular to digitize 3D objects in the world of computer science. For complex objects, to construct and to render their 3D models are often very difficult. Hence, in our digital museum project working together with National Palace Museum and National Museum of History, we adopt object movie approach [1, 2] for digitizing antiques.

Object movie which is first proposed by Apple Computer in QTVR (QiuckTime VR) [1] is an image-based rendering approach [3–6] for 3D object representation. An object movie is generated by capturing a set of 2D images at different perspectives around the real object. Figure 1 illustrates the image components of an object movie to represent a Wienie Bear. During the process of capturing an object movie, Wienie Bear is fixed and located at center, and the camera location is around Wienie Bear by controlling pan and tilt angles, denoted as θ and ϕ, respectively. Instead of constructing a 3D model, the photos captured at different viewpoints of the Wienie Bear are collected to be an object movie for representing it. The more photos for the object we have, the more precise the corresponding representation is.

Some companies, for example, Kaidan and Texnai, provide efficient equipments to acquire object movies in an easy way. Object movie is appropriate to represent real and complex objects for its photo-realistic view effect and for its ease of acquisition. Figure 2 shows some examples of antiques that are included in our object movie database.

The goal of this paper is to present our efforts in developing an efficient approach for retrieving desired object in an object movie database. Consider a simple scenario. A sightseer is interested in an antique when he visits a museum. He can take one or more photos of the antique at arbitrary viewpoints using his handheld device and retrieve related guiding information from the Digital Museum. Object movie is

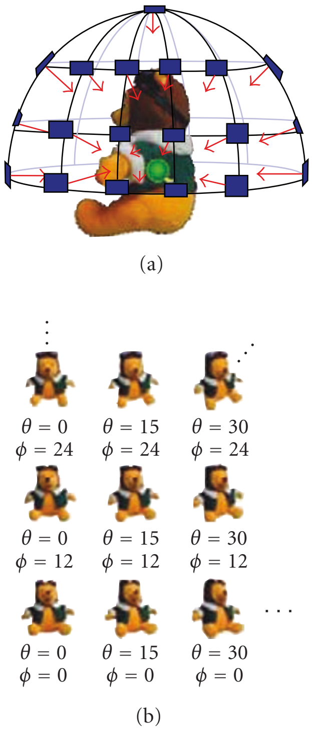


FIGURE 1: The image components of an object movie. The left shows the camera locations around Wienie Bear, and the right shows some captured images and their corresponding angles.

a good representation for building the digital museum because it provides realistic descriptions of antiques but does not require 3D model construction. Many related works of 3D model retrieval which are described in Section 2 have been published. However, to our best knowledge, we do not find any literatures that work on content-based object movie retrieval.

In this paper, we mainly focus on three issues: (i) the representation of an object movie, (ii) matching and ranking for object movies, and (iii) relevance feedbacks for improving the retrieval results. A design of two-layer feature descriptor, comprising dense and condensed, is used for representing an object movie. The goal of the *dense descriptor* is to describe an object movie as precise as possible while the *condense descriptor* is its compact representation. Based on the two-layer feature descriptor, we define dissimilarity measure between object movies for matching and ranking. The basic idea of the proposed dissimilarity measure between the query and target object movie is that if two objects are similar, the observation of them from most viewpoints will be also similar. Moreover, we apply relevance feedbacks approach to iteratively improving the retrieval results.

The rests of this paper are organized as follows. In Section 2, we review some related literatures for 3D object retrieval. Our proposed two-layer feature descriptor for object movie representation is described in Section 3. Next, the dissimilarity measure between object movies is designed in Section 4. In Section 5, we present our design of relevance feedbacks for improving object movie retrieval. Related experiments are presented in Section 6 for showing the efficacy of our proposed approach. Finally, Section 7 gives some conclusions of this work and possible directions of future works.

2. RELATED WORK

Content-based approach has been widely studied for multimedia information retrieval, such as images, videos, and 3D objects. The goal of content-based approach is to retrieve the desired information based on the contents of query. Many researches of content-based image retrieval have been published [7–9]. Here, we focus on related works of 3D object/model retrieval based on content-based approach.

In [10], Chen et al. proposed the LightField Descriptor to represent 3D models and defined a visual similarity-based 3D model retrieval system. The LightField Descriptor is defined as features of images rendered from vertices of dodecahedron over a hemisphere. Note that Chen et al. used a huge database containing more than 10,000 3D models collected from internet in their experiments.

Funkhouser et al. proposed a new shape-based search method [11]. They presented a web-based search engine system that supports queries based on 3D sketches, 2D sketches, 3D models, and text keywords.

Shilane et al. described the Princeton Shape Benchmark (PSB) [12] which is a publicly available database of 3D geometric models collected from internet. The benchmarking dataset provides two levels of semantic labels for each 3D model. Note that we adopt PSB as our test data in our experiment.

Zhang and Chen presented a general approach for indexing and retrieval of 3D models aided by active learning [13]. Relevance feedback is involved in the system and combined with active learning to provide better user-adaptive retrieval results.

Atmosukarto et al. proposed an approach of combining the feature types for 3D model retrieval and relevance feedbacks [14]. It performs the query processing based on known relevant and irrelevant objects of the query and computes the similarity to an object in the database using precomputed rankings of the objects instead of computing in high-dimensional feature spaces.

Cyr and Kimia presented an aspect-graph approach to 3D object recognition [15]. They measured the similarity between two views by a 2D shape metric of similarity which measures the distance between the projected and segmented shapes of the 3D object.

Selinger and Nelson proposed an appearance-based approach to recognizing objects by using multiple 2D views [16]. They investigated the performance gain by combining the results of a single view object recognition system with imagery obtained from multiple fixed cameras. Their approach also addresses performance in cluttered scenes with varying degrees of information about relative camera pose.

Mahmoudi and Daoudi presented a method based on the characteristic views of 3D objects [17]. They defined seven


(a) (b) (c)

FIGURE 2: Some examples of museum antiques included in our object movie database.

characteristic views which are determined by the eigenvector of analysis of the covariance matrix related to the 3D object.

3. REPRESENTATION FOR AN OBJECT MOVIE

3.1. Sampling in an object movie

Since an object movie is the collection of images captured from the 3D object at different perspectives, the construction of an object movie can be considered the sampling of 2D viewpoints of the corresponding object. Figure 3 shows our basic idea to represent an object movie. Ideally, we can have an object movie consisting of infinite views, that is, infinite images, to represent a 3D object. By extracting the feature vector for each image, the representation of an object movie forms a manifold in the feature space. However, it is impossible to take infinite images of a 3D object. We can simply regard the construction of an object movie as a sampling of some feature points in the corresponding manifold in the feature space. In general, the denser the sampling of the manifold we have, the more accurate the object movie is represented. Note that the sampling idea for an object movie is independent of the selection of visual features.

Figure 4 illustrates the sampling of the manifold corresponding to the object movie which contains 2D images around Wienie Bear at a fixed tilt angle. This example plots a closed curve which represents the object movie in the feature space and illustrates the relationship between the feature points and the viewpoints for the object movie. Since drawing a manifold in high dimensional space is difficult, we simply chose 2D features which comprise the average hue for the vertical axis and the first component of Fourier descriptor of the centroid distance for the horizontal axis. The curve approximates the manifold of the object movie using the sampling feature points.

3.2. Dense and condensed descriptors

In estimating the manifold of an object movie, the denser the sampling of feature points can perform, the better representation, but it also implies high computational complexity in object movie matching and retrieval. Our idea is to design dense and condensed descriptors which provide different densities in the sampling of the manifold to balance the accuracy and computational complexity.

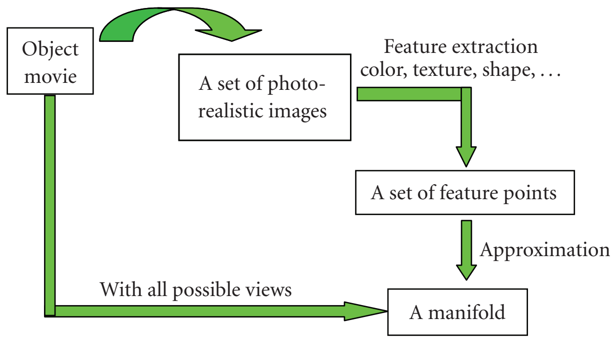


FIGURE 3: Representation of an object movie.

Both the dense and condensed descriptors are the collection of sampling feature points of the manifold in the feature space. The dense descriptor is designed to sample feature points as many as possible, hence it consists of feature vectors that are extracted from all 2D images of an object movie. Suppose that an object movie O is the set $\{I_i\}$, $i = 1$ to M, where each I_i is an image, that is, a viewpoint, of the object, and M is the number of images captured from O. Let F_i be the feature vector extracted from image I_i, then we define the feature set $\{F_i\}$, $i = 1$ to M as the dense descriptor of O.

The main idea of designing the condensed descriptor is to choose the key aspects of all viewpoints of the object movies. We adopt K-means clustering algorithm to divide the dense descriptor $\{F_i\}$ into K clusters, denoted as $\{C_i\}$, $i = 1$ to K. For each cluster C_i, choose a feature point $R_i \in C_i$ such that R_i is the closest point to the centroid of C_i. Then, we define the set $\{R_i\}$, $i = 1$ to K as the condensed descriptor of O. The condensed descriptor is the set of more representative feature points sampled from the manifold for an object movie. In general, K-means clustering is sensitive to initial seeds. That is to say, the condensed descriptor may be different if we perform K-means clustering again. This is not very critical because the goal of the condensed descriptors is to roughly sample the dense descriptor.

To represent and compare the query and a target object movie in the database using the dense and condensed descriptors, there are four possible cases: (i) both the query and the target using the dense descriptor, (ii) the query using the dense descriptor and the target using the condensed descriptor, (iii) the query using the condensed descriptor

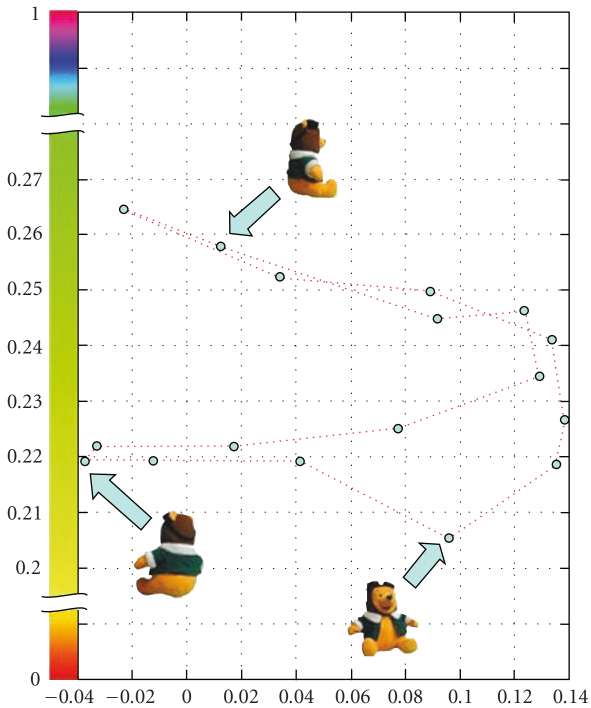


Figure 4: A curve representing an object movie in the feature space. Each feature point corresponds to a view of the object.

and the target using the dense descriptor, and (iv) both the query and the target using the condensed descriptor. Case (i) would be simple but inefficient, case (ii) does not make sense in efficient reason, and case (iv) would be too coarse in object movie representation. Since the representation of object movies in the database can be done offline, we would like to represent them as precise as possible. Therefore, dense descriptor is preferred for the object movies in the database. In contrast, a query from the user is supposed to be processed quickly, so condensed descriptor is preferred for the query. Hence, we adopt case (iii) in order to balance both accuracy and speed issues in retrieval.

3.3. *Visual features*

Our proposed descriptors, either dense or condensed, are independent of the selection of visual features. In this work, we adopt color moments [18] for color feature, Fourier descriptor of centroid distances [19], and Zernike moments [20, 21] for shape features.

Color moments

Stricker and Orengo [18] used the statistical moments of color channels to overcome the quantization effects in the color histogram. Let x_i be the value of pixel x in ith color component, and let N be the pixel number of the image. The first- and second-order color moments of an image are defined as

$$\begin{aligned} \mathrm{CM} &= \left(\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3\right), \\ \text{where } \mu_i &= \frac{1}{N}\sum_{x=1}^{N} x_i, \ \sigma_i = \frac{1}{N}\sum_{x=1}^{N}\left(x_i - \mu_i\right)^2. \end{aligned} \tag{1}$$

Thus, color moments are six dimensional. In our work, we adopt Lab color space for this feature.

Fourier descriptor of centroid distance

The centroid distance function [19] is expressed by the distances between the boundary points and the centroid of the shape. The centroid distance function can be written as

$$r(t) = \left[\left(x(t) - x_c\right)^2 + \left(y(t) - y_c\right)^2\right]^{1/2}, \tag{2}$$

where $x(t)$ and $y(t)$ denote the horizontal and vertical coordinates, respectively, of the sampling point on the shape contour at time t, and (x_c, y_c) is the coordinate of the centroid of the shape. Then, the sequence of centroid distances is applied to Fourier transformation as the Fourier descriptor of centroid distances. There are some invariant characteristics in Fourier descriptor of centroid distances, including rotation, scaling, and change of start point from an original contour.

In our implementation, we take 128 sampling points on the shape contour for each image. That is to say, a sequence of centroid distances will contain 128 numbers. Then, we derive Fourier transformation for getting 63D vectors of the Fourier descriptor of centroid distances. Finally, we reduce the dimension of this feature vector to 5D by PCA (principal component analysis).

Zernike moments

Zernike moments are a class of orthogonal moments and have been shown effective in terms of image representation [21]. The Zernike polynomials $V_{nm}(x, y)$ [20, 21] are a set of complex orthogonal polynomials defined over the interior of a unit circle. Projecting the image function onto the basis set of Zernike polynomials, the Zernike moments, $\{|A_{nm}|\}_{n,m}$, of order n with repetition m are defined as

$$\begin{aligned} A_{nm} &= \frac{n+1}{\pi}\sum_{x}\sum_{y} f(x, y) V_{nm}(x, y), \\ &\qquad \text{where } x^2 + y^2 \leq 1, \end{aligned} \tag{3}$$

$|A_{nm}|$ is the magnitude of the projections of image function, and Zernike moments are a set of the projecting magnitudes. Zernike moments are rotation invariant for an image. Similarly, we reduce the dimension of Zernike moments to 5D by PCA.

4. OBJECT MOVIE MATCHING AND RETRIEVAL

In our work, we handled two types of queries: a set of viewpoints (single or multiple viewpoints) of an object and an entire object movie. Both two query formats can be considered a set of viewpoints of an object.

Let Q be the query, either a set of viewpoints of an object or an entire object movie, and let O be candidate object movies in the database. In this work, our idea is to regard the query Q as a mask or a template such that we can compute the matching scores to candidate object movies in the database by fitting the query mask or the query template. We take the condensed descriptor for Q and dense descriptor for O. Then, Q and O can be represented as $\{R_i^Q\}_{i=1}^k$ and $\{F_j^O\}_{j=1}^n$, respectively, where R_i^Q and F_j^O are image features mentioned in Section 3.2. Then, we define the dissimilarity measure between Q and O as

$$d(Q,O) = \sum_{i=1}^{K} p_i \cdot d(R_i^Q, O) = \sum_{i=1}^{K} p_i \cdot \min_j d(R_i^Q, F_j^O), \quad (4)$$

where $d(R_i^Q, O)$ is the shortest Euclidean distance from R_i^Q to all feature points $\{F_j^O\}_{j=1}^n$, and the weight p_i is the size percentage of the cluster C_i^Q to which R_i^Q belongs. Thus, the dissimilarity measure $d(Q,O)$ is a weighted summation of each dissimilarity $d(R_i^Q, O)$.

Since we choose three types of visual features to represent the 2D images, we then revise (4) for cooperating with different types of features by weighted summation of dissimilarities in individual feature spaces:

$$\begin{aligned} d(Q,O) &= \sum_c w_c \cdot d_c(Q,O) \\ &= \sum_c w_c \sum_{i=1}^{k} p_i \cdot \min_j d_c(R_i^Q, F_j^O), \end{aligned} \quad (5)$$

where $d_c(R_i^Q, F_j^O)$ means the Euclidean distance from R_i^Q to F_j^O in the feature space c, and w_c is the important weight of the feature c in computing the dissimilarity measure. We set the equal weights in the initial query, that is, $w_c = 1/C$, where C is the number of visual features used in the retrieval.

5. RELEVANCE FEEDBACK

The performance of content-based image retrieval being unsatisfactory for many practical applications is mainly due to the gap between the high-level semantic concepts and the low-level visual features. Unfortunately, the contents in images for general purpose retrieval are much subjective. Relevance feedback (RF) is a query modification technique that attempts to capture the user's precise needs through iterative feedback and query refinement [8]. There have been many tasks of content-based image retrieval for applying relevance feedbacks [22–24]. Moreover, Zhang and Chen adopted active learning for determining which objects should be hidden and annotated [13]. Atmosukarto et al. tune the weights of combining feature types by use of positive and negative examples of relevance feedbacks [14].

We summarize the standard process of relevance feedback in information retrieval as follows.

(1) The first query is issued.
(2) The system computes the matching ranks of all data in the database and reports some of them.
(3) The user specifies some relevant (or positive) and irrelevant (or negative) data from the results of step 2.
(4) Go to step 2 to get the retrieval results of the next iteration according to relevant and irrelevant data until the user do not continue the retrieval.

We design a relevance feedback that reweights features of the dissimilarity function by use of users' positive feedbacks. Here, we rewrite (5) by attaching a notation t, for describing feedback iterations:

$$d(Q,O) = \sum_c w_{ct} \cdot d_{ct}(Q,O), \quad (6)$$

where $d_{ct}(Q,O)$ denotes the dissimilarity measure between object movie Q and O in feature space c at iteration t, and w_{ct} means its weight.

Next, we introduce how to decide the weight of a feature c according to users' feedbacks. We compute the scatter measure, defined as the accumulated dissimilarities among pairs of feedbacks within feature space c at the iteration t, as

$$s(c,t) = \sum_i \sum_{j \neq i} d_c(O_{ti}, O_{tj}), \quad (7)$$

where both O_{ti} and O_{tj} are feedback examples at the iteration t. Thus, we express the importance of feature c as the inverse of summation of scatter measures computed in past iterations:

$$f_c = \left(\sum_{i=1}^{t} s(c,i) \right)^{-1}. \quad (8)$$

Based on the importance of features, f_c, we then reassign weights of features using the weighting function shown below, where W_t is a matrix which comprises the weights w_{ct} associated with feature c at tth iteration

$$W_{t+1} = (1-\alpha) \cdot W_t + \alpha \cdot M_t, \quad (9)$$

$$M_{t,k} = \begin{cases} 1, & \text{if } k = \arg\min_c f_c \\ 0, & \text{otherwise} \end{cases}, \quad k = 1, \dots, C. \quad (10)$$

In these two equations, C is the number of features, W and M are $C \times 1$ matrices, and α is the learning rate. Note that $M_{tk} = 1$ indicates that feature type k is the most significant to represent the relevant examples at tth iteration of the relevance feedbacks. Also, we set α to 0.3 in our implementation.

6. EXPERIMENTAL RESULTS

6.1. Data set

We have a collection of object movies of real antiques that is from our Digital Museum project working together with National Palace Museum and National Museum of History. However, we also need a large enough object movie databases and their ground truth labeling for the quantitative evaluation of our proposed system. We do not have hundreds of

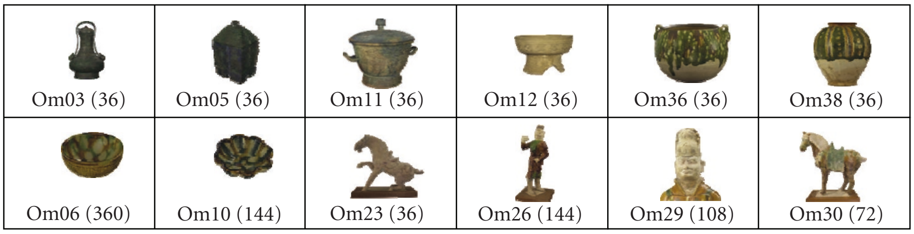


Figure 5: OMDB1: the index and number of images for some objects.

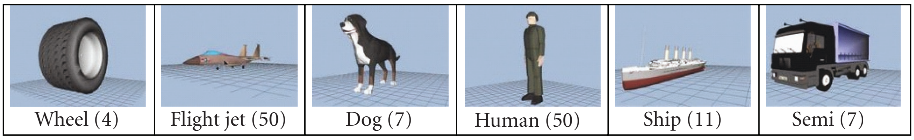


Figure 6: OMDB2: the semantic name and the object number for some classes of base classification.

object movies to perform the retrieval experiments. Hence, instead of using real object movie directly, we collected many 3D geometric models and transformed them to other object movie databases for simulation.

The first database used in the experiments, called OMDB1 and listed in Figure 5, contains 38 object movies of real antiques. The numeric in the image caption is the number of 2D images taken from the 3D object. All color images in these object movies were physically captured from the antiques.

The second database, OMDB2, is the collection of simulated object movies taken from the benchmarking dataset Princeton Shape Benchmark [12]. We captured 2D images by changing pan, ϕ, and tilt, φ, angles by 15° for each object movie. Thus, there are $(360/15) \times (180/15 + 1) = 312$ images for each object movie. This dataset contains 907 objects, and two classification levels, base and coarse, are involved to be the ground truth labeling in our experiments. All data are classified as 44 and 92 classes in the base and coarse levels, respectively. Some examples of classes are listed in Figure 6.

Because the object movies in the OMDB1 are captured from real artifacts, all 2D images are colorful and textural. We adopted color moments, Fourier descriptor of centroid distances, and Zernike moments as the features ($C = 3$ in (6)) for representing images of object movies. However, all object movies in OMDB2 are not rendered really, we only chose shapes features, Fourier descriptor of centroid distance, and Zernike moments as the features ($C = 2$ in (6)).

6.2. Evaluation

We used the precision/recall curve to evaluate the performance of our system on the three object movie database. Note that $precision = B/A'$ and $recall = B/A$, where A' is the number of retrieved object movies, B is the number of retrieved relevant ones, and A is the number of all relevant ones in the database. Next, we design three kinds of experiments for measuring the performance of our approach at different perspectives.

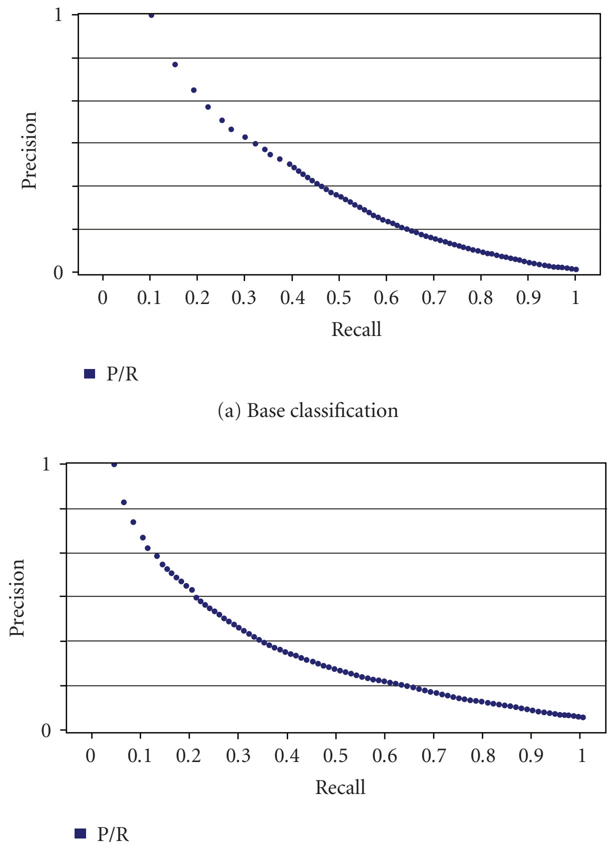
Table 1: Comparison of results with queries comprising 1, 3, 5, and 10 views in OMDB1.

Feature	1 view	3 views	5 views	10 views
Fourier descriptor	74.4%	92.6%	95.4%	97%
Zernike moments	81.6%	95%	97.2%	97.4%
Color moments	94.8%	98.8%	99.8%	99.8%
Combination	99%	99.8%	100%	100%

OMDB1 without relevance feedbacks

This experiment aims at showing the efficacy of our approach in the dataset of real objects. OMDB1 contains a small size of object movies of real antiques, so it is not proper to apply the relevance feedback approach in this dataset. We only considered the retrieval results of the first query in OMDB1. We took some views, rather than the entire, of an object movie as the query. The retrieved object is relevant only if it is the same as the query object. That is similar to object recognition.

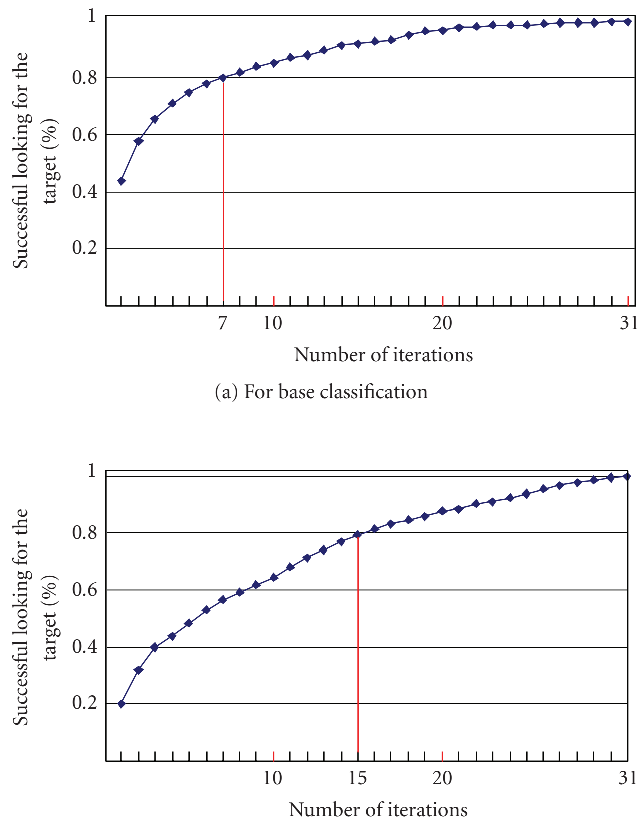
We randomly chose v views from an object movie to be the query, where v is set as 1, 3, 5, and 10. These taken query views were removed from OMDB1 in each test. Table 1 shows the average precisions of queries (by repeating the random selection of a query 500 times to compute the average) using different number of views. These results show that among the three features we used, color moment has better performance in this experiment, and combining these features can even provide excellent results approaching 99% of retrieval that target can be found on the first rank using only one view.



(a) Base classification

(b) Coarse classification

Figure 7: The average precision-recall curves of base and coarse classifications in OMDB2.



(a) For base classification

(b) For coarse classification

Figure 8: Evaluation for target search: percentage of successful search with respect to the number of iterations.

OMDB2 without relevance feedbacks

This experiment aims at presenting the quantitative measure of the performance for our proposed approach. Two levels of semantic labels comprising base and coarse are assigned in OMDB2, hence more semantic concepts are involved in this dataset. We employed an entire object movie as the query for observing the retrieval results at different semantic levels. Figure 7 shows the average precisions/recalls for OMDB2, where Figures 7(a) and 7(b) are the performances of choosing the ground truth labeling base and coarse classifications, respectively.

OMDB2 with relevance feedbacks

We adopt target search [25] for evaluating the experiment of relevance feedback. In our experiment, the procedure of target search for a test is summarized as follows.

(1) The system randomly chooses a target from database, and let G be the class of the target.

(2) The system randomly chooses an object from the class G as the initial query object.

(3) Execute query process and examine the retrieves. If the target is in the top H retrieval results, the retrieval is stop; otherwise go to step 4. In our implementation, we set the H as 30.

(4) Pick the object movies in class G within top H results as relevant ones.

(5) Apply the process of relevance feedbacks by use of relevant object movies. Then go to step 3.

Output: the number of iterations is used for reaching the target.

Based on base and coarse levels individually, 900 object movies are randomly taken as targets from the database. For each target, we apply target search five times for computing the average number of iterations. Figure 8(a) shows the average number of iterations of target search based on base classification, and Figure 8(b) shows that based on coarse classification.

For the successful rate 80% of the target search shown in Figures 8(a) and 8(b), 7 and 15 iterations are computed for the base and coarse classes, respectively. That is to say, the results for the base classes are better than that for the coarse classes. The reason is that objects in the coarse classes are more various. The positive examples for a query may be also very different in the coarse classes. For example, both object movies with bikes and with trucks are relevant in the base and coarse levels, respectively, for an object movie with a bike. The feedbacks with bike can indicate more precise and correct information than those with truck.

7. CONCLUSION

The main contribution of our paper is to propose a method for retrieving object movies based on their contents. We propose dense and condensed descriptors to sample the manifold associated with an object movie. We also define the dissimilarity measure between object movies and design a scheme of relevance feedback for improving the retrieval results. Our experimental results have shown the potential of this approach. Two future tasks are needed to extend this work. The first is to apply negative examples in relevance feedbacks to improve the retrieval results. The other task is to employ state of the art of content-based multimedia retrieval and relevance feedback to the object movie retrieval.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Economic Affairs, Taiwan, under Grant 95-EC-17-A-02-S1-032 and by the Excellent Research Projects of National Taiwan University under Grant 95R0062-AE00-02.

REFERENCES

[1] S. E. Chen, "QuickTime VR—an image-based approach to virtual environment navigation," in *Proceedings of the 22nd Annual ACM Conference on Computer Graphics and Interactive Techniques*, pp. 29–38, Los Angeles, Calif, USA, August 1995.

[2] Y.-P. Hung, C.-S. Chen, Y.-P. Tsai, and S.-W. Lin, "Augmenting panoramas with object movies by generating novel views with disparity-based view morphing," *Journal of Visualization and Computer Animation*, vol. 13, no. 4, pp. 237–247, 2002.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics (SIGGRAPH '96)*, pp. 43–54, New Orleans, La, USA, August 1996.

[4] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics (SIGGRAPH '96)*, pp. 31–42, New Orleans, La, USA, August 1996.

[5] L. McMillan and G. Bishop, "Plenoptic modeling: an image-based rendering system," in *Proceedings of the 22nd Annual Conference on Computer Graphics (SIGGRAPH '95)*, pp. 39–46, Los Angeles, Calif, USA, August 1995.

[6] C. Zhang and T. Chen, "A survey on image-based rendering—representation, sampling and compression," *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[7] V. Castelli and L. D. Bergman, *Image Databases: Search and Retrieval of Digital Imagery*, John Wiley & Sons, New York, NY, USA, 2002.

[8] R. Datta, J. Li, and J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age," in *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR '05)*, pp. 253–262, Singapore, November 2005.

[9] R. Zhang, Z. Zhang, M. Li, W.-Y. Ma, and H.-J. Zhang, "A probabilistic semantic model for image annotation and multimodal image retrieval," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 1, pp. 846–851, Beijing, China, October 2005.

[10] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," *Computer Graphics Forum*, vol. 22, no. 3, pp. 223–232, 2003.

[11] T. Funkhouser, P. Min, M. Kazhdan, et al., "A search engine for 3D models," *ACM Transactions on Graphics*, vol. 22, no. 1, pp. 83–105, 2003.

[12] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape Benchmark," in *Proceedings of Shape Modeling International (SMI '04)*, pp. 167–178, Genova, Italy, June 2004.

[13] C. Zhang and T. Chen, "An active learning framework for content-based information retrieval," *IEEE Transactions on Multimedia*, vol. 4, no. 2, pp. 260–268, 2002.

[14] I. Atmosukarto, W. K. Leow, and Z. Huang, "Feature combination and relevance feedback for 3D model retrieval," in *Proceedings of the 11th International Multimedia Modelling Conference (MMM '05)*, pp. 334–339, Melbourne, Australia, January 2005.

[15] C. M. Cyr and B. B. Kimia, "3D object recognition using shape similarity-based aspect graph," in *Proceedings of the 8th International Conference on Computer Vision (ICCV '01)*, vol. 1, pp. 254–261, Vancouver, BC, USA, July 2001.

[16] A. Selinger and R. C. Nelson, "Appearance-based object recognition using multiple views," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 905–911, Kauai, Hawaii, USA, December 2001.

[17] S. Mahmoudi and M. Daoudi, "3D models retrieval by using characteristic views," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 2, pp. 457–460, Quebec, Canada, August 2002.

[18] M. A. Stricker and M. Orengo, "Similarity of color images," in *Storage and Retrieval for Image and Video Databases III*, vol. 2420 of *Proceedings of SPIE*, pp. 381–392, San Jose, Calif, USA, February 1995.

[19] D. S. Zhang and G. Lu, "A comparative study of Fourier descriptors for shape representation and retrieval," in *Proceedings of the 5th Asian Conference on Computer Vision (ACCV '02)*, pp. 646–651, Melbourne, Australia, January 2002.

[20] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

[21] H. Hse and A. R. Newton, "Sketched symbol recognition using Zernike moments," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 367–370, Cambridge, UK, August 2004.

[22] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 815–818, Santa Barbara, Calif, USA, October 1997.

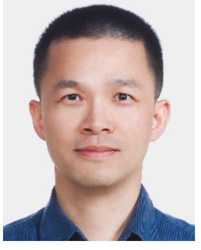
[23] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, 2003.

[24] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: a comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, 2003.

[25] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos, "PicHunter: Bayesian relevance feedback for image retrieval," in *Proceedings of the 13th International Conference on Pattern Recognition (ICPR '96)*, vol. 3, pp. 361–369, Vienna, Austria, August 1996.

Cheng-Chieh Chiang received a B.S. degree in applied mathematics from Tatung University, Taipei, Taiwan, in 1991, and an M.S. degree in computer science from National Chiao Tung University, HsinChu, Taiwan, in 1993. He is currently working toward the Ph.D. degree in Department of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan. His research interests include multimedia information indexing and retrieval, pattern recognition, machine learning, and computer vision.

Li-Wei Chan received the B.S. degree in computer science in 2002 from Fu Jen Catholic University, Taiwan, and the M.S. degree in computer science in 2004 from National Taiwan University. He is currently taking Ph.D. program in Graduate Institute of Networking and Multimedia, National Taiwan University. His research interests are interactive user interface, indoor localization, machine learning, and pattern recognition.



Yi-Ping Hung received his B.S. degree in electrical engineering from the National Taiwan University in 1982. He received an M.S. degree from the Division of Engineering, an M.S. degree from the Division of Applied Mathematics, and a Ph.D. degree from the Division of Engineering, all at Brown University, in 1987, 1988, and 1990, respectively. He is currently a Professor in the Graduate Institute of Networking and Multimedia, and in the Department of Computer Science and Information Engineering, both at the National Taiwan University. From 1990 to 2002, he was with the Institute of Information Science, Academia Sinica, Taiwan, where he became a tenured research fellow in 1997 and is now an adjunct research fellow. He served as a deputy director of the Institute of Information Science from 1996 to 1997, and received the Young Researcher Publication Award from Academia Sinica in 1997. He has served as the program cochairs of ACCV '00 and ICAT '00, as the workshop cochair of ICCV '03, and as a member in the editorial board of the International Journal of Computer Vision since 2004. His current research interests include computer vision, pattern recognition, image processing, virtual reality, multimedia, and human-computer interaction.

Greg C. Lee received a B.S. degree from Louisiana State University in 1985 and M.S. and Ph.D. degrees from Michigan State University in 1988 and 1992, respectively, all in Computer Science. Since 1992, he has been with the National Taiwan Normal University where he is currently a Professor at the Department of Computer Science and Information Engineering. His research interests are in the areas of image processing, video processing, computer vision, and computer science education. Dr. Lee is a Member of IEEE and ACM.