CONTENT-BASED TRADEMARK RETRIEVAL WITH FUZZY CLUSTERING AND RELEVANCE FEEDBACK
使用乏晰分群及相關回饋技術之內容導向式商標讀取系統

Peng-Yeng Yin
Department of Computer Science, Ming Chuan University
5 Teh-Ming Rd., Taoyuan 333, Taiwan
pyyin@mcu.edu.tw

尹邦嚴
銘傳大學資訊科學系
桃園縣龜山鄉德明路 5 號

Abstract *Since the number of registered trademarks is increasing rapidly, the job of identifying infringement of similar trademarks by human inspection becomes laborious and time consuming. To deal with the problem, we propose an automatic content-based trademark retrieval method. The proposed method automatically selects appropriate features based on feature selection principles to discriminate trademarks. The database trademarks are softly clustered into classes using a fuzzy approach to increase the retrieval speed. The user can submit a query through trademark examples to get a list of database trademarks ordered by similarity ranks. The query results can be iteratively refined by the feedback presented by the user until the trademarks of interest are retrieved. Experiments are conducted on a trademark database containing 1000 images and the retrieval results are very encouraging.*

Keywords. Content-based image retrieval; Fuzzy c-means clustering; Relevance feedback; Trademark

摘要　由於註冊商標的數目成長快速,由人工方式檢測侵權的工作變得相當費時,因此我們提出一個自動化商標讀取系統,我們的方法根據特徵挑選原則來決定特徵,並將商標影像以乏晰權重分入數群以提高讀取速率,使用者可提出一個查詢商標影像,以獲得一組最相似的資料庫商標,查詢結果可以利用使用者的回饋不斷繼續改進,直到使用者滿意為止,我們在一個含有 1000 張商標影像的資料庫上做測試,實驗結果令人滿意.

關鍵字　內容導向式影像讀取,乏晰 c 均值分群法,相關回饋,商標

1. Introduction

Due to the increasing demands of managing pictorial data such as art galleries, remote sensing, criminal investigation, medical image archiving and trademark management, the development of efficient image storage and retrieval systems becomes extremely important. Traditional database systems are not capable of manipulating pictorial data since they manage data in alphanumeric form. Recently, many content-based image retrieval systems have emerged to satisfy the needs [1-2].

In this paper, we investigate the problem of trademark management. Trademarks identify the goods or services of companies. Once the trademark is registered, it is assumed to receive the protection from the infringement of the design. Applications of registration are now processed manually to ensure that no registered trademarks are potentially infringed. The task is laborious and time-consuming since the number of registered trademarks is hundreds of thousands in many countries. As a result, the development of an automatic trademark management system becomes crucial. There are some content-based trademark retrieval systems proposed in the last decade. Kato et al. [3] first developed an automatic trademark management system using statistical features of mesh image and segmental element. Kim and Kim [4] used the most salient feature (MSF) based on Zernike moments to describe the global view of trademarks. Cortelazzo et al. [5] represented the inclusion relationships between contours of trademark objects by a contours-tree with chain-code strings. Then the shape similarity is measured by calculating the string distance using dynamic programming technique. Peng and Chen [6] encoded the sequential internal angles of the closed contours into angle-code strings instead of chain-code strings, so the properties of translation, scaling, and rotation invariance can be preserved. However, they did not consider the relative orientation between the sequential internal angles, which can detect considerable variations along the shape contour even if the angle-code strings of the two objects are similar. Lu and Tsai [7] used statistical and shape features to describe the contents of trademarks more detailedly and proposed a coarse-to-fine classification scheme to speed up the retrieval process. Nevertheless, some of their selected features are too complicated and should be extracted manually. Jain and Vailaya [8] showed the better performances of using a combination of color and shape features than using either of the individuals for trademark retrieval. Mehtre et al. also confirmed this idea by combining color and shape features [9]. Jain and Vailaya [10] proposed a two-stage trademark retrieval system based on multiple models of shape features. In the first stage, the edge directions and the invariant moments are used to prune the database to plausible candidates. Then in the second stage, a deformable template matching process is used to discard spurious ones. More recently, an interesting topic called relevance feedback [11] has attracted many researchers. Since the importance of each selected feature may be different for various users, one possible solution is to let the individual user indicate the positive/negative instances on the retrieval result, meaning that he or she is most satisfied/unsatisfied with the specified instances. The system would be then able to adapt the contributing weight for each feature according to the feedback and yield more appropriate retrieval result for that user.

In this paper, we propose an approach for content-based trademark retrieval. The approach selects a set of appropriate features automatically according to feature selection principles. To increase the retrieval speed, the training images are clustered to several classes using a fuzzy approach. The matching of the query image is made to the subset of database images which are softly clustered to the corresponding classes. We also propose a new scheme for the implementation of the relevance feedback technique. The experiments on a database of 1000 trademarks are conducted and the results show that the proposed approach is feasible.

2. The Proposed Approach

The block diagram of the proposed approach is shown in Fig. 1. It can be separated into two stages. In the training stage, 1000 training trademark images are preprocessed. Then the features of all training images are extracted, and these images are clustered into several classes using a fuzzy approach according to their feature values. In the query stage, users can submit a query by a trademark example to get a list of database trademarks ranked by their similarity scores to the query image. The details of the proposed approach are presented in the following subsections.

2.1 The training stage

2.1.1 Preprocessing

We have constructed a database of 1000 trademark images. Most of them are registered trademarks and are scanned from the publications of Taiwan Central Bureau of Standards. First each image is thresholded to binary version. Then the minimal enclosing rectangle (MER) of the trademark objects is properly scaled to put into a blank image of 128*128 pixels such that the aspect ratio of the MER is unchanged. Thus the training images are invariant to scaling and translation variations after the preprocessing step.

2.1.2 Feature extraction

We select a set of appropriate features according to two principles. First, the feature value spectrum associated with each feature should carry as much information as possible to discriminate the trademarks. For each considered feature candidate, we calculate the corresponding feature values for all training trademarks. The feature values are quantized into 100 bins for normalization. Let h_i be the number of training trademarks which have feature values falling into bin i, we evaluate the entropy of the feature value spectrum as

$$E = -\sum_{i=1}^{100} p_i \log p_i \,, \quad (1)$$

where

$$p_i = h_i \Big/ \sum_{i=1}^{100} h_i \,. \quad (2)$$

The larger the entropy value, the more informative the feature. The candidate features with entropy less than a threshold are considered to be less informative and should be discarded.

The measure of the feature value entropy testifies the discriminating ability of each feature independently, but two entropy-qualified features may be redundant if they are very correlated. Hence we measure the correlation between every pair of entropy-qualified features to remove the redundant features. The correlation between feature j and feature r is given by

$$C(j,r)=\frac{\left|\sum_{i=1}^{n}(f_{ij}-m_j)(f_{ir}-m_r)\right|}{n\sigma_j\sigma_r}, \quad (3)$$

where n is the number of trademarks in the database, f_{ij} and f_{ir} are the values of features j and r for the i-th trademark, m_j, σ_j, m_r and σ_r are the mean values and the standard deviations of features j and r. The value of correlation is between 0 and 1. Higher correlation means the two corresponding features are more correlated. Two features are too correlated if their correlation value is larger than a specified threshold, and the one with less entropy value is discarded.

Finally, we determine 7 features described as follows.

<1> Area. The number of black pixels.

<2> Isolation. The number of connected components.

<3> Deviation. The distance deviation from the mass centroid (center of all black pixels) to the geometric image center.

<4> Symmetry. The smaller value of $\frac{H}{V}$ and $\frac{V}{H}$, where H and V are the numbers of black runs along the horizontal and vertical directions, respectively.

<5> Centralization. The number of black pixels resident in the interior of the circle which is drawn at the image center with a radius of one fifth of the image width.

<6> Complexity. First, the boundary of each connected component of the trademark image is approximated by line segments using Ray and Ray's polygonal approximation technique [12]. Then the feature is evaluated as the sum of the numbers of line segments of all approximating polygons in the image.

<7> 2-level contour representation strings. First the angles between the intersections of successive line segments along boundaries of the approximating polygon are calculated and quantized into 8 bins of 22.5° each, so the angles are coded in the range [0, 7] as shown in Fig. 2(a). The sequence of the internal angle codes is not unique depending on the starting angle. We can normalize the codes by treating the sequence as a circular list and choosing the starting angle such that the resulting sequence of numbers forms an integer of minimum magnitude [13]. We define this sequence of angle codes as the first level contour representation string. For instance, Fig. 2(b) shows the approximating polygon of an object and the quantity of its internal angles. The corresponding first level contour representation string is 26446 by choosing 60° as the starting angle. Secondly, to calculate the relative orientation between sequential internal angles, the direction of the left wing of each internal angle (i.e., the preceding line segment of each internal angle if we trace the contour clockwise) according to the sequence of the first level string is evaluated. The direction is encoded to one of the direction codes as shown in Fig. 2(c). Therefore, the second level contour representation string of Fig. 2(b) is 75420. The details of the matching method using the 2-level contour representation strings will be described in Subsection 2.2.2.

2.1.3 Database building

Since image database usually has a large volume of pictorial data, an efficient indexing method is crucial. Previous researchers have used the hard clustering technique to partition the database into several non-overlapping clusters. However, the similar images of the query image are not necessarily associated with the same cluster. As a result, we propose to use the fuzzy c-means algorithm (FCMA) [14] where each element can belong to several clusters simultaneously with various membership scales (soft clustering). Hence, the images closest to the query instead of the images of the closest cluster to the query will be retrieved.

2.2 The query stage

The user can submit a query through a sample image, the proposed system will output a list of database images ranked by their similarity scores to the query image. The query stage can be divided into two steps. First, the clusters to which the query image belongs with a membership value larger than a threshold are determined. Then the database images of the clusters should be matched with the query image. The details are presented as follows.

2.2.1 Cluster membership of the query image

Let Q be the query image, z_i be the centroid of the i-th cluster. We first compute the initial fuzzy cluster membership that the query image Q belongs to the i-th cluster as follows.

$$y(Q,z_i)=\begin{cases}\dfrac{1}{\sum_{j=1}^{c}\sqrt{(\frac{d_i}{d_j})^{\frac{2}{m-1}}}}, & \text{if } d_i \neq 0,\\ 1 & \text{, if } d_i = 0,\end{cases} \quad (4)$$

where m is the fuzziness level, d_i denotes the Euclidean distance between Q and z_i, and c is the number of clusters. The larger the value of m, the fuzzier the cluster membership. Experiments show that the setting m = 1.2 is appropriate for our database. The final cluster membership of Q is then computed by

$$\omega(Q,z_i)=\begin{cases}y(Q,z_i), & \text{if } y(Q,z_i)\geq\tau\\ 0 & \text{, if } y(Q,z_i)<\tau\end{cases} \quad (5)$$

where τ is the cluster membership threshold. Thus, all the database images in the clusters to which Q belongs with a membership value that is equal to or larger than τ should be matched with Q.

2.2.2 Matching

Let database image P be determined to match with Q, the matching process is consisting of two parts. The first part is concerned with the similarity regarding to each of the first six features of P and Q and is given by

$$S_i(P,Q)=\frac{w_i}{1+\dfrac{|p_i-q_i|}{\sigma_i}}, \quad 1\leq i\leq 6 \quad (6)$$

where p_i and q_i are the i-th feature values of P and Q, σ_i is the standard deviation of the i-th feature values for all database images, and w_i is the contributing weight of the i-th feature. The similarity measure is based on the normalized distance metric for each feature and integrates various features by different weights. The values of these weights are determined empirically.

The second part of matching is related to the similarity between the 2-level contour representation strings of P and Q. Before we describe the details, some terminologies are first introduced. A string is a subsequence of another string if it is obtained by removing some elements (not necessarily adjacent) from the latter string. Let $SUB(S)$ denote the set of all subsequences of string S, then a string S_3 is called the common subsequence (CS) of strings S_1 and S_2 if $S_3 \in SUB(S_1) \cap SUB(S_2)$. The longest CS of S_1 and S_2, denoted by $LCS(S_1, S_2)$, represents the similarity degree of S_1 and S_2 in one aspect. There exist efficient algorithms in the literature [15] to compute the longest CS. Let A_i and B_j be the first level contour representation strings of the i-th and j-th connected components of P and Q, respectively, we now define the similarity measure between the first level contour representation strings as

$$S^{1st-level}(A_i,B_j)=\frac{2|LCS(A_i,B_j)|}{|A_i|+|B_j|}, \quad (7)$$

where | • | denotes the length of the corresponding string. Eq. (7) measures the similarity between shape contours in terms of the number of common sequential internal angles. However, two shapes could vary significantly even if they have high similarity score evaluated from Eq. (7). Fig. 3 shows such an example. The two objects look quite different although they have four common internal angles in the longest *CS*. Therefore, the relative orientation between sequential common angles should be taken into account. As illustrated in Fig. 3, the relative orientation between angle θ_1 and angle θ_3, which is evaluated as the angle between the extending lines of the left wings of angle θ_1 and angle θ_3, reflects part of the difference between the two objects. Since we have encoded the directions of the left wings of all internal angles in the second level contour representation string, the relative orientation can be evaluated as the difference between the two direction codes. Let d_1 and d_2 be the direction codes of the left wings of two internal angles of a shape, we define the relative orientation of d_1 and d_2 as

$$\nabla(d_1,d_2)=\begin{cases}|d_1-d_2-8| & \text{if } d_1-d_2>4\\ |d_1-d_2+8| & \text{if } d_1-d_2<-3\\ |d_1-d_2| & \text{otherwise.}\end{cases} \quad (8)$$

The value of ∇ is in the range of [0, 4] where one unit represents 45° difference in the direction. For example, ∇ = 2 means the relative orientation of the two angles is 90°, ∇ = 0 indicates the two angles are facing the same orientation, and ∇ = 4 indicates the two angles are in opposite orientations.

Let P and Q have s and t connected components and (A_i, D_i) and (B_j, E_j) be the two-tuple vectors of the 2-level contour representation strings of the i-th and j-th connected components of P and Q, respectively. Let C be the $LCS(A_i, B_j)$. Assume $A_i = a_1^i a_2^i \cdots a_m^i$,

$B_j = b_1^j b_2^j \cdots b_n^j$, and $C = c_1 c_2 \cdots c_k$. There exist two monotonically increasing functions $p(\cdot)$ and $q(\cdot)$ such that $c_h = a^i_{p(h)}$ and $c_h = b^j_{q(h)}$ for $1 \le h \le k$. Assume $D_i = d_1^i d_2^i \cdots d_m^i$ and $E_j = e_1^j e_2^j \cdots e_n^j$, the subsequences of D_i and E_j mapped by the $LCS(A_i, B_j)$ are $d^i_{p(1)} d^i_{p(2)} \cdots d^i_{p(k)}$ and $e^j_{q(1)} e^j_{q(2)} \cdots e^j_{q(k)}$, respectively. We define the similarity measure between the second level contour representation strings as

$$S^{2nd-level}(D_i, E_j) = \frac{2 \sum_{1 \le a < b \le k} \left(1 - \frac{1}{4}\left|\nabla(d^i_{p(a)}, d^i_{p(b)}) - \nabla(e^j_{q(a)}, e^j_{q(b)})\right|\right)}{k(k-1)}. \quad (9)$$

The value of this measure is in range [0, 1] and is inversely proportional to the cumulated difference between the relative orientations of pairs of sequential common internal angles. Thus the similarity measure of the 2-level contour representation strings of P and Q can be derived based on Eqs. (7) and (9) as

$$S_7(P, Q) = w_7 \max_{\substack{1 \le i \le s \\ 1 \le j \le t}} \left(S^{1st-level}(A_i, B_j) + S^{2nd-level}(D_i, E_j)\right) \quad (10)$$

where w_7 is the weight for the matching score of the 2-level contour representation strings.

Finally, combining the two parts of the similarity measures defined in Eqs. (6) and (10), we propose the similarity measure for comparing a database image P to a query image Q as

$$S(P, Q) = \sum_{i=1}^{7} S_i(P, Q). \quad (11)$$

2.3 Relevance feedback

The proposed system will output the retrieval result in an ordered list of images according to their similarity scores evaluated from Eq. (11). As mentioned previously, the relevance feedback technique allows the user to present a feedback by indicating the retrieved images as positive or negative instances to refine the query. In this subsection, we propose a new scheme to implement the relevance feedback technique. Let $CI(Q)$ be the set of database images to which the query image Q was determined to compare according to its cluster membership, m_i and σ_i be the mean and the standard deviation of the set of similarity scores, $\{S_i(P_j, Q) \text{ for every } P_j \in CI(Q)\}$, with respect to the i-th feature. Assume P^+ be one of the positive instances presented by the user, we evaluate the influence of P^+ on the i-th feature as

$$x_i(P^+) = \left(S_i(P^+, Q) - m_i\right) \big/ \sigma_i. \quad (12)$$

The increasing rate α_i^+ of the weight w_i caused by P^+ is obviously related to $x_i(P^+)$. We propose to use an exponential function to determine α_i^+ as

$$\alpha_i^+ = \begin{cases} \min(0.004 e^{5x_i(P^+)} - 0.004, 1), & \text{if } x_i(P^+) \ge 0 \\ \max(0.004 - 0.004 e^{-5x_i(P^+)}, -1), & \text{if } x_i(P^+) < 0 \end{cases} \quad (13)$$

The relation between α_i^+ and $x_i(P^+)$ is illustrated in Fig. 4. It is seen that α_i^+ grows exponentially with the increment of $x_i(P^+)$, and vice versa. The value of α_i^+ is bounded by [-1, 1]. The case of the negative instance P^- can be derived similarly and we get

$$\alpha_i^- = \begin{cases} \max(0.004 - 0.004 e^{5x_i(P^-)}, -1), & \text{if } x_i(P^-) \ge 0 \\ \min(0.004 e^{-5x_i(P^-)} - 0.004, 1), & \text{if } x_i(P^-) < 0 \end{cases} \quad (14)$$

We adapt the contributing weight of each feature by the feedback of all positive and negative instances as

$$w_i = w_i\left(1 + \sum_{\forall P^+} \alpha_i^+ + \sum_{\forall P^-} \alpha_i^-\right), \quad \text{for } 1 \le i \le 7. \quad (15)$$

If any weight becomes negative, it is set to zero and will not be adapted in the following feedback iterations. Finally, the weights are normalized by

$$w_i = w_i \Big/ \sum_{j=1}^{7} w_j, \quad \text{for } 1 \le i \le 7. \quad (16)$$

3. Experimental Results

At present, the proposed system has been trained with 1000 trademark images and these images were softly partitioned into five clusters using FCMA with fuzziness level set to 1.2. The cluster memberships of the training images are thresholded by 0.26, and each training image belongs to 1.7 clusters on average. The optimal parameter setting is based on the comparative performance analysis.

Fig. 5 shows three examples of the retrieval results by the proposed system. The query image is shown on the top. The query image is matched with a subset of database images determined by its cluster membership. The retrieved results are output in the order of their similarity ranks. Suppose the system displays a list of t most similar images to the query image according to the ranks of the similarity measure. Let the number of images which are really similar (judged by experts) to the query image be n, and r images of them are retrieved in the output list by the system. The retrieval efficiency, η_t, is defined by [9]

$$\eta_t = \begin{cases} \dfrac{r}{t}, & \text{if } n > t \\ \dfrac{r}{n}, & \text{if } n \le t \end{cases} \tag{17}$$

To assess the retrieval performance of the proposed system, we randomly choose 50 images from the database and present each of them as the query image. The retrieval results are judged by 10 different experts to see if the retrieved images are similar to the query image. Fig. 6 illustrates the average retrieval efficiency of the proposed system with various lengths of the output list.

To determine the best number of clusters to partition the test database, the comparative performances of the retrieval efficiency and the retrieval time are analyzed. The experiments are conducted with the number of clusters set to 1, 3, 5, 7, and 9, and a list of 10 most similar images is output for each query image. Fig. 7 shows that the retrieval efficiency drops from 90% to 50% as the number of clusters increases, but the saving of retrieval time is also dramatic. It is appropriate to partition our database into 5 clusters where satisfied retrieval efficiency is obtained and the retrieval time is very short. This analysis is very important when we extend our approach to very large databases that may contain millions of images.

To evaluate the significance of the proposed relevance feedback technique, the previous random-selected query images are presented as query images again, and the retrieved results are further refined for two iterations through the proposed relevance feedback method. The average retrieval efficiency of the original result and the other two refinements are listed in Table 1. It is seen that the retrieval efficiency increases along with the number of iterations of relevance feedback process and the improving ratio of the efficiency contributed by the first feedback iteration (16%) is larger than that by the second iteration (7%).

4. Conclusions

In this paper, we have presented an approach for content-based trademark retrieval. The proposed method selects appropriate features based on the principles of information entropy and correlation. All the trademarks in the database are softly clustered into 5 classes using FCMA to increase the retrieval speed. The experiments are conducted on a database of 1000 trademarks. The user can improve the query result by presenting a feedback of positive and negative instances. The limitation of the proposed method is the semantic meanings of the images are unknown while human beings usually interpret images by their semantics.

References

[1] M. Flickner and H. Sawhney, Query by image and video content: the QBIC system, IEEE Comput. 23 (1995) 23-32.

[2] A. Pentland, R. W. Picard and S. Sclaroff, Photobook: content-based manipulation of image databases, International Journal of Computer Vision 18 (1996) 233-254.

[3] T. Kato and K. Fujimura, TRADEMARK: multimedia image database system with intelligent human interface, Systems and Computers in Japan 21 (1990) 33-46.

[4] Y. S. Kim and W. Y. Kim, Content-based trademark retrieval system using a visually salient feature, Image and Vision Computing 16 (1998) 931-939.

[5] G. Cortelazzo, G. A. Mian, G. Vezzi and P. Zamperoni, Trademark shapes description by string-matching techniques, Pattern Recog. 27 (1994) 1005-1018.

[6] Peng and Chen, Trademark shape recognition using closed contours, Pattern Recog. Letters 18 (1997) 791-803.

[7] W. C. Lu and W. H. Tsai, Trademark image recognition using statistical and shape features, in: Proceedings of IPPR Conf. on Computer Vision, Graphics and Image Processing, Taiwan, (1994) 26-33.

[8] A. K. Jain and A. Vailaya, Image retrieval

using color and shape, Pattern Recog. 29 (1996) 1233-1244.

[9] B. M. Mehtre, M. S. Kankanhalli and W. F. Lee, Content-based image retrieval using a composite color – shape approach, Information Processing and Management 35 (1998) 109-120.

[10] A. K. Jain and A. Vailaya, Shape-based retrieval: a case study with trademark image databases, Pattern Recog. 31 (1998) 1369-1390.

[11] B. Bhanu, J. Peng and S. Qing, Learning feature relevance and similarity metrics in image database, Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, CA (1998), 14-18.

[12] B. K. Ray and K. S. Ray, Determination of optimal polygon from digital curve using L_1 norm, Pattern Recog. 26 (1993) 505-509.

[13] R. C. Gonzalez and P. Wintz, Digital Image Processing, Addison Wesley, Reading, 1987.

[14] M. S. Kamel and S. Z. Selim, A thresholded fuzzy c-means algorithm for semi-fuzzy clustering, Pattern Recog. 24 (1991) 825-833.

[15] S. Kuo and G. R. Cross, A two-step string-matching procedure, Pattern Recog. 24 (1991) 711-716.

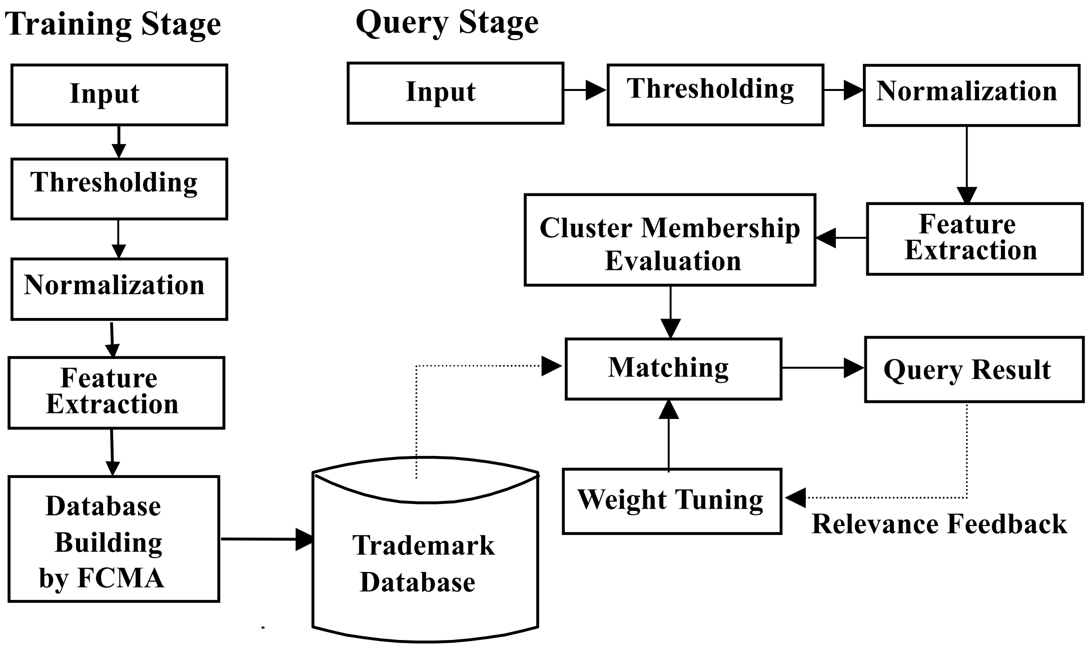

Fig. 1 The block diagram of the proposed approach.

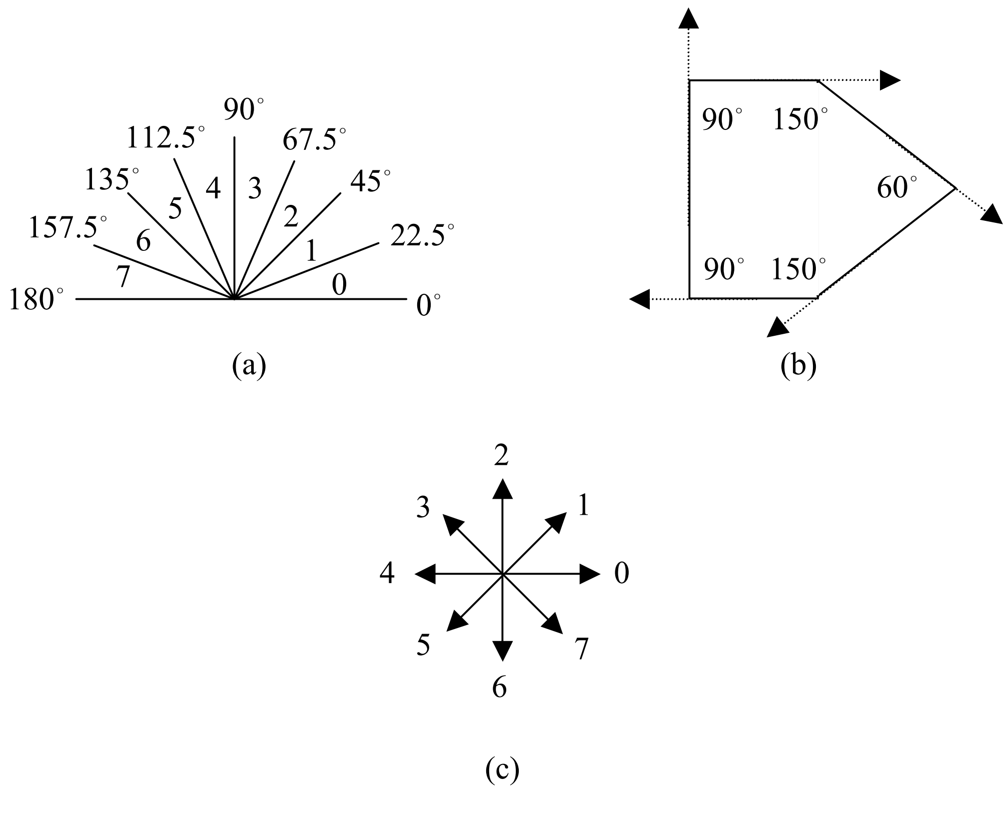

Fig. 2 The illustration of the 2-level contour representation strings.

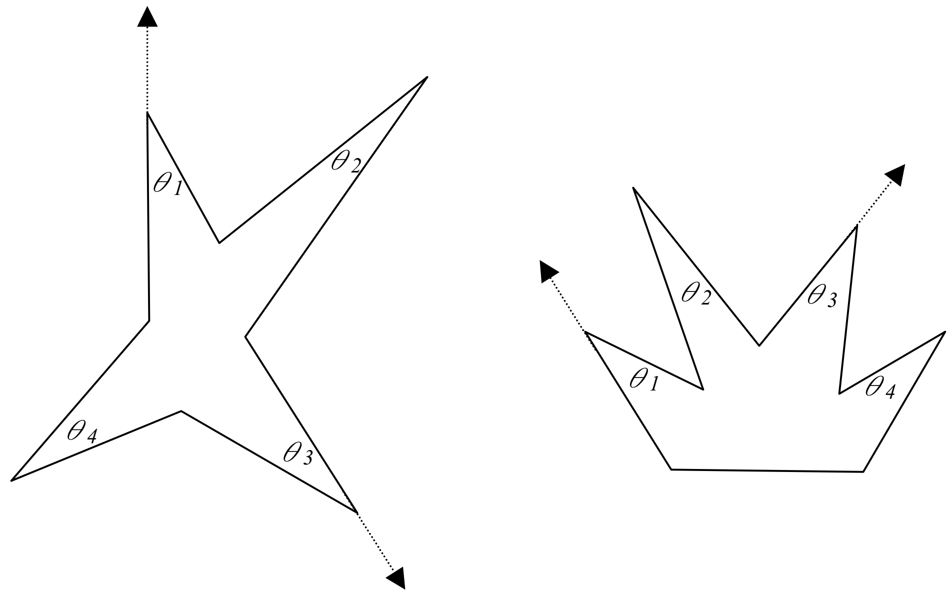

Fig. 3 Two objects have four common internal angles; however, the relative orientations between angles are quite different.

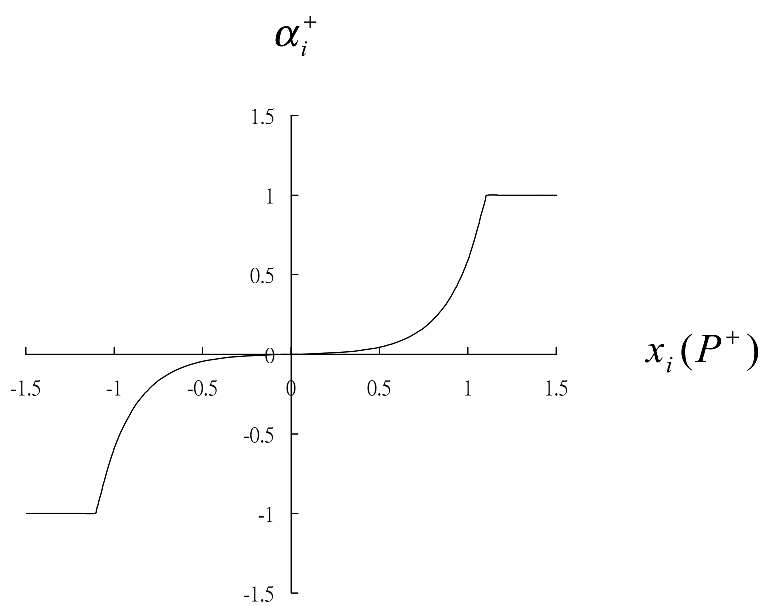

Fig. 4 The relation between α_i^+ and $x_i(P^+)$.

Fig. 5 Examples of the retrieval results.

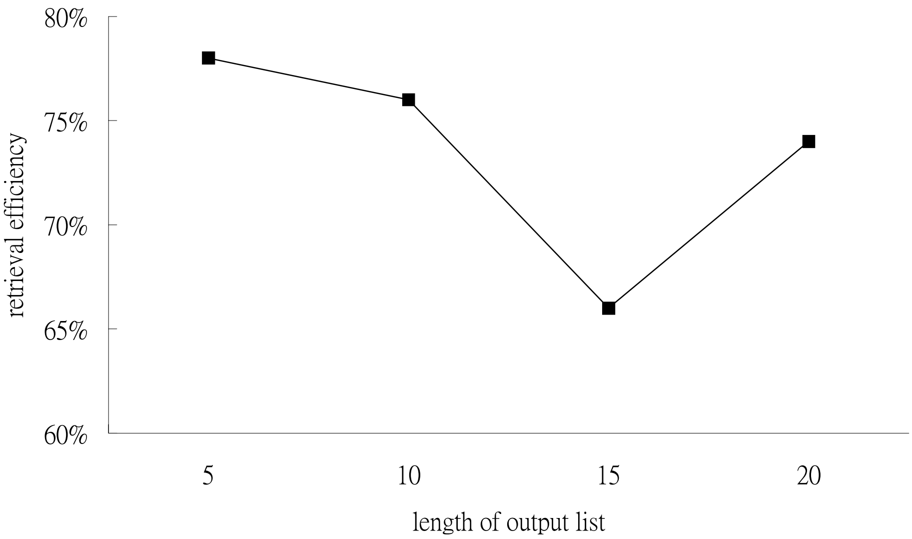

Fig. 6 The average retrieval efficiency with various lengths of the output list.

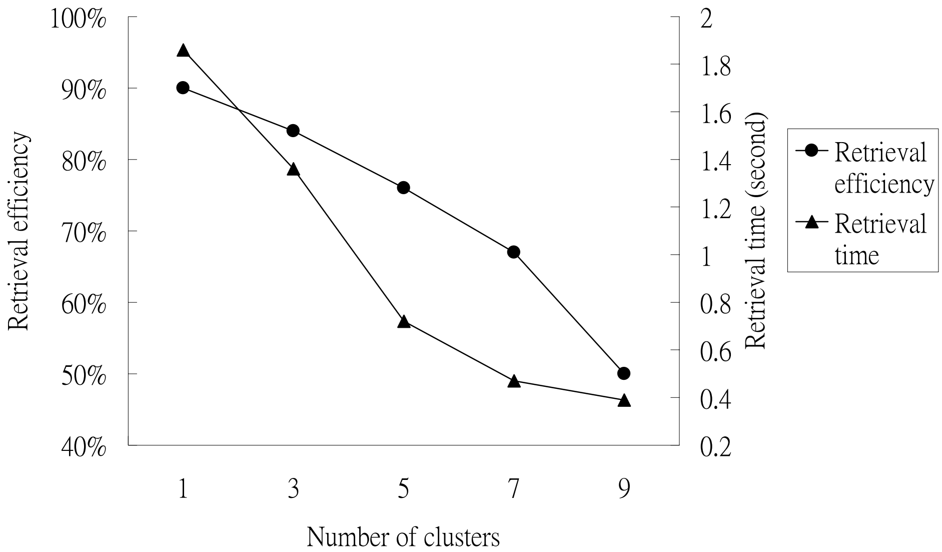

Fig. 7 The retrieval efficiency and the retrieval time with respect to different number of clusters.

Table 1 The average retrieval efficiency and the improving ratio using the proposed relevance feedback technique.

	Original result	1st feedback iteration	2nd feedback iteration
Retrieval efficiency	76%	88%	94%
Improving ratio		16%	7%