A Performance-Driven Rotational Invariant Image Retrieval System*

YEN-TAI LAI, CHI-CHOU KAO†, TZU-CHIANG TAI AND WEN-CHUN YEH
Department of Electrical Engineering
National Cheng Kung University
Tainan, 701 Taiwan
†Department of Computer Science and Information Engineering
National University of Tainan
Tainan, 700 Taiwan

The color feature could be well represented in a feature and shape is a distinguishable feature which can make coarse discriminations quickly. Based on those reasons, we present an efficient image retrieval system to consider simultaneously accuracy and performance by combining color moment with shape contexts in this paper.

The shape rotation has significant influence on the similarity measure and retrieval performance. Another important contribution of the paper is to present an efficient algorithm for rotation invariance. In the algorithm, the rotated shape contexts together are first clustered and then label each cluster so that the shape contexts in each cluster have the same label. Using the histogram of label frequencies can quickly and efficiently search for similar or rotational shapes. The experimental results have shown that our system is effective and has better retrieval performance than the existing systems.

Keywords: image retrieval, performance, rotation-invariance, color, shape context

1. INTRODUCTION

With the amount of digital multimedia data growing, the efficient information retrieval issue is raised. Most of the early researches developed content-based image retrieval. The content-based image retrieval system was focused on extracting effective global visual features i.e. low-level features, such as color, texture, and shape to compare images [1-4]. However, the retrieval performance is still far from user’s expectations. It is well known that the most challenging problems of content-based image retrieval are the semantic gap between low-level image representations and high-level human concepts. Two approaches can be used to reduce the gap. The first approach is extracting region-based features to represent the user’s perception. The region-based image retrieval [5-8] applies image segmentation to decompose an image into several regions and extract features of these regions. Therefore, each image can be represented by a set of regions, which contain low-level features and importance weights. According to the weight, the system determines which regions are interested. Compared to the case where global image features are considered, the region-based image retrieval system is close to the perception of human visual system. The second approach taken to reduce the gap is using relevance feedback. This approach employs an online learning scheme to improve the
retrieval extraction performance by applying positive and negative examples according to the user’s subjective perception [9, 10]. Since every user’s subjectivity is different, the system can become personality by relevance feedback.

To retrieve similar images, we should compute the similarity between images. Since objects can be described by various features, we evaluate the similarity by the image properties. However, the complexity is increased by the number of features. It is important to utilize a few features that are representative. Shape is a distinguishable feature which can make coarse discriminations quickly. Therefore, adding shape matching can increase the performance of an image retrieval system [11]. Besides, each color feature could be well represented in a feature space. Accordingly, we will focus on shape matching for region-based image retrieval in the paper. In our system, we use color moment and shape contexts to describe a region. Although using multiple features increases complexity, the proposed system can filter out quickly candidate images by establishing a color codebook and a shape codebook, respectively.

The shape context is one of the important shape features in the image retrieval system. In past, many shape analysis techniques have been proposed [12-18]. For compact shape representation and matching of 2D contours, well known methods include Fourier Descriptors (FD) [12], Curvature Scale Space (CSS) [13], and approaches based on the classical Hausdorff distance [14, 15]. Another class of contour matching algorithms includes methods which attempt to solve the correspondence problem between two shapes and measure the similarity based on this assignment. The optimal match between contour segments is found using Dynamic Programming [16]. In [17], the shape context is assigned to each contour point in order to solve the correspondence problem. The vector quantization technique is proposed for shape context to quickly prune a search for similar shapes [18].

The important contribution of the paper is to present an efficient algorithm for rotation invariance. The algorithm takes into account the shape context for rotated objects when performing clustering, and assign the same shape labels to a shape and its rotation. Thus, our system can find similar shapes in the database for a query image no matter they are rotations with each other. The experimental results have shown that our system is effective and has better retrieval performance than the existing systems.

The remainder of this paper is organized as follows. In section 2, we outline the proposed image retrieval system and formulate the matching problem for rotated object. An efficient rotation invariance algorithm is presented in section 3 and the experimental results are summarized in section 4. Finally, we conclude in section 5.

2. THE SYSTEM OVERVIEW

The flowchart of the proposed image retrieval system is shown in Fig. 1. At the beginning, a user inputs a query image into the system. In the system, all images in the database and the query image are segmented into homogenous regions. We employed the segmentation algorithm, which is based on local homogeneity analysis [19]. Next, color and shape features are extracted from all regions. For the database, images with similar color and shape regions are clustered to form a color codebook or a shape codebook, respectively. Each region in the query image will be encoded into a color codeword and
a shape codeword. The system will compare all the color and shape codebooks with the color and shape codewords to find the best matching images. If the user is satisfied with the results, the search is finished. Otherwise, the users choose correct images from the results to the system. When system receives the correct images, it will learn human concept and retrieve more accurate images in the next time.

Image similarity between the query and database can be computed by utilizing extracted features. After segmentation and extraction, each regional color feature is encoded to a codeword. According to the features, an image can be represented as compact and sparse representations. The similarity of two images is decided by the distance of their representation. The shorter is the distance, the more similar the images. To speed up the retrieval, database is structured as a codebook to filter quickly out candidate images. All images in the database are segmented into homogeneous regions and then the regional features in those images are extracted. Similar features will be clustered together and encoded as a codeword. The database can be considered as a codebook. The first column of a codebook is a codeword and the second column stores \( k \) codewords similar to the first column one and sorted by the similarity, where \( k \) is a natural number. The third column collects a list of images that have a region corresponding to the first column codeword. According to those codewords, we can obtain a set of images and compute the distances between the query image and the set.

If the matching results are not satisfied, the system will learn the importance of a region and perform a new query representation. A region \( R \) and an image \( I \) are defined to
be similar if at least one region of $I$ is similar to $R$. Intuitively, the larger the region frequency is, the more important this region is. To reflect the distinguishing ability of a region, a measure of inverse image frequency (IIF) for region $R_i$ is defined as:

$$IIF(R_i) = \log \left( \frac{N}{\sum_{j=1}^{N}(R_i, I_j)} \right).$$

Therefore, the importance of a region (RI) is:

$$RI(R_i) = \frac{RF(R_i) \times IIF(R_i)}{\sum_{j=1}^{n}(RF(R_j) \times IIF(R_j))}.$$ 

Obviously, the importance of a region is its region frequency weighted by the inverse image frequency, and normalized over all regions in the image such that the sum of importance of all regions is equal to 1.

After computing the RIs of all images, a new query representation is performed. Assume that there are $n$ images $\{I_1, \ldots, I_m, \ldots, I_n\}$, with $\{I_1, \ldots, I_m\}$ being the prior ones and $\{I_{m+1}, \ldots, I_n\}$ being the new ones. Let the sparse representation of a positive image $I_k$ is: $I_k = (w_{k,1}, \ldots, w_{k,N})$. Let the new query $I_{new}$ is: $I_{new} = (w_{new,1}, \ldots, w_{new,N})$. Let $\alpha \geq 0$ be the factor that controls the importance of the prior images. The smaller is $\alpha$, the less important is the prior positive regions. The new query is defined to be

$$I_{new} = \beta \left( \sum_{k=1}^{m} I_k + \sum_{k=m+1}^{n} I_k \right),$$

where $\beta$ serves as a normalization factor to make $I_{new}$ satisfy the constraint: $\sum_{j=1}^{N} w_{new,j} = 1$.

Using new query to retrieve images is closing the user’s concept, but new query may contain more codewords. For reducing the complexity, the codewords of the new query with the corresponding weights less than a threshold $\xi$ are pruned directly.

The color feature could be well represented in a feature and shape is a distinguishable feature which can make coarse discriminations quickly. However, the image retrieval system, which only uses color matching method, has worse matching performance. Moreover, the image retrieval system, which only uses shape matching method, can generate more unsatisfying results such that the number of feedbacks is largely increased. Any one of the two systems can decrease the total performance of the system. The most important feature in the proposed relevance feedback image retrieval system is to consider simultaneously accuracy and performance by combining color moment with shape contexts. Additionally, we also present an efficient algorithm for rotation invariance in the shape matching step because the shape rotation has significant influence on the similarity measure and retrieval performance. The rotation invariance algorithm will be described in detail in the next section.
3. A ROTATIONAL-INVARIANT ALGORITHM

In this section, we first describe retrieval by shape contexts to formulate the rotated object matching problem. Next, the rotation invariance algorithm based on the $k$-means clustering method [20] is proposed.

3.1 Problem Formulation

Shape context is a rich shape descriptor for measuring shape similarity. The shape context describes the coarse arrangement of the shape with respect to a given point on the shape and this point is not required to be any special point such as maxima of curvature or inflection point. The shape context analysis begins by taking $n$ samples from the edge elements on the shape. Then, a shape is represented by a discrete set of points sampled from its contours. These points can be obtained as locations of edge pixels as found by an edge detector, giving us a set $P = \{p_1, p_2, \ldots, p_n\}, p_i \in \mathbb{R}^2$, of $n$ points. When we consider the set of vectors originating from a point to all other sample points on a shape, these $n-1$ vectors express the configuration of the entire shape relative to the reference point. One compact way to capture this information is the distribution of the relative positions of the remaining $n-1$ points in a spatial histogram. Concretely, for a point $p_i$ on the shape, we compute a coarse histogram $h_i$ of the relative coordinates of the remaining $n-1$ points,

$$h_i^k = \#\{q \neq p_i: \{q - p_i\} \in \text{bin}(m)\}.$$

This histogram is defined to be the shape context of $p_i$. Thus, each shape can obtain $n$ shape contexts. To make the descriptor more sensitive to positions of nearby sample points than to those of points farther away, we use bins that are uniform in log-polar space. It should be noticed that in the absence of background clutter, the shape context of a point on a shape is made scale invariant by normalizing all radial distances by the mean distance $\alpha$ between the $n^2$ point pairs in the shape.

We observe that shape contexts will be different for different points on a single shape $S$; however, corresponding (homologous) points on similar shapes $S$ and $S'$ will tend to have similar shape contexts. As illustrated in Figs. 2 (a)-(c), sampled edge points for one shape are at left and the shape context of the marked point is at right. Each shape context is a log-polar histogram of the coordinates of the rest of the points measured using the marked point as the origin; dark pixels represent larger values. Figs. 2 (a) and (b) have similar shape contexts since their marked points are corresponding points on similar shapes. Figs. 2 (a) and (c) have different shape contexts although the two marked points are on the same shape.

By using shape contexts as shape descriptor, we can quickly determine which shapes in the database are similar to the query shape. The basic idea is using vector quantization on the shape contexts. With $|S|$ known shapes, and shape contexts computed at $n$ sample points on these shapes, the full set of shape contexts for the known shapes consists of $|S| \times n d$-dimensional vectors where $d$ is the total number of bins in a shape context histogram. A compression technique for dealing with such a large amount of data is vector quantization. Vector quantization involves clustering the vectors and then representing each vector by the index of the cluster that it belongs to. These indexes are called shape labels.
To represent each shape with shape labels, all of the shape contexts from the known set are first transformed to $d$-dimensional vectors and considered as points in a $d$-dimensional space. These vectors are called shape context vectors. Then the $k$-means clustering can be performed to obtain $k$ clusters and label each group by an integer in $\{1, 2, \ldots, k\}$. Each $d$-dimensional shape context vector is quantized to its nearest clusters and replaced by its shape label. A known shape is then simplified into a histogram of shape label frequencies. We have reduced each collection of $n$ shape contexts ($d$ bin histograms) to a single histogram with $k$ bins.

Since the corresponding points on similar shapes have similar shape contexts, similar shapes will have similar shape contexts. If we use a shape label to replace a shape context, then similar shapes will have similar histogram of shape label frequencies. This property is used to match a query shape. The same vector quantization and histogram creation operation is performed on the shape contexts from the query shape. We then find nearest neighbors in the space of histogram of shape label frequencies.

However, if a shape is rotated, the distribution of shape labels will be different. The corresponding points (which are marked) have different shape contexts for one shape and its rotation by 180° and are replaced with different shape labels. Obviously, this causes dissimilar histograms of shape label frequencies between the shape and its rotation. If we perform image retrieval by using the set of shape labels directly, the rotated object will not be found. We need an efficient algorithm to solve this problem.

3.2 The Rotation Invariance Algorithm

To achieve rotation invariance in terms of traditional shape context, it is essential to have the same collection of shape labels for a shape and its rotations. We propose a rotation invariance algorithm based on the $k$-means clustering method to cluster together not only the shape context vectors of similar shapes but also the corresponding shape context vectors of original shape and its rotations. Therefore, a shape label represents a kind of distribution based on relative frame, rather than the absolute frame in traditional shape context.

For two corresponding points on one shape and its rotation, we can find that one
shape context vector can be derived by circularly shifting another shape context vector. By using this kind of shifting on a shape context vector, there are different shape context vectors as many as the number of bins for $\theta$. Each shape context vector corresponds to one rotation that can be realized with respect to the original shape in the log-polar space. Fig. 3 illustrates an example where one shape is the rotation of the other shape by $180^\circ$. The log-polar space is uniformly divided into 12 bins for $\theta$ and 5 bins for log $r$. We can observe that the shape context vector in Fig. 3 (a) can be derived when we circularly shift the shape context vector in Fig. 3 (b) by 30 positions. ($180^\circ$ is equal to 5 bins for $\theta$ and rotating through one bin for $\theta$ is equal to 6 positions.)

From the discussion above, we consider that all shape context vectors that can be derived by circular shift represents the corresponding points for one shape and all of its rotations. Therefore, these shape context vectors should be clustered together and replaced with the same shape label. For this purpose, we propose the rotation invariance algorithm as follows.

First, the absolute distance and representative shape context vector for 2 shape context vector in our log-polar space:

For two shape context vectors $X = (x_1, x_2, \ldots, x_{60})$ and $Y = (y_1, y_2, \ldots, y_{60})$, the absolute distance of $X$ and $Y$ is denoted as $D_{XY}$ and its computation is

$$D_{XY} = \min_i \|X - R_i\|, 1 \leq i \leq 12,$$

where:

- $R_1 = Y = (y_1, y_2, y_3, y_4, y_5, \ldots, y_{60})$
- $R_2 = (y_6, y_7, y_8, \ldots, y_{60}, y_1, y_2, y_3, y_4, y_5)$
- $R_3 = (y_{11}, y_{12}, y_{13}, \ldots, y_5, y_6, y_7, y_8, y_9, y_{10})$
- $\vdots$
- $R_{12} = (y_{56}, y_{57}, y_{58}, y_{59}, y_{60}, \ldots, y_{53}, y_{54}, y_{55})$.

The representative shape context vector is denoted as $Y'$ and defined as

$$Y' = \arg\min_{R_i} \|X - R_i\|, 1 \leq i \leq 12.$$
their shape context. The smaller the absolute distance, the more similar the two shape contexts are. We can see the absolute distance will be zero for two shape context vectors when their shape contexts are derived from corresponding points of a shape and its rotation.

The flowchart of our rotation invariance algorithm is shown in Fig. 4. The input to this algorithm is all training shape context vectors that we want to classify and the number of cluster, $k$. The output is $k$ clusters and $k$ centroids for each cluster. The rotation invariance algorithm groups the vectors with small absolute distance together and use the representative vector to compute the centroid of each group. Suppose there are $N$ shape contexts in the database. The steps for the rotation invariance algorithm are further described as follows.

**Step 1:** When there are $N$ shape contexts in the database, the user decide the number of clusters, $k$.

**Step 2:** Select $k$ shape contexts arbitrarily, and consider them as the centroid of each cluster.

**Step 3:** Use our proposed method to calculate the distance between each shape context and the centroid. That is, fixing the vector of centroid and rotating the vector of shape context by a multiple of a constant. Compute the minimum distance between rotated vectors and centroid, and record the rotated vector which is minimal.

**Step 4:** Group shape contexts according to the minimal distance. Compute the shape context which is closest to the centroid and include it into the group.
**Step 5:** If a shape context switches its cluster in step 4, update the centroid of the cluster:

$$C_i = \sum_{Y' \in \text{group}_i} Y', \ 1 \leq i \leq k.$$

**Step 6:** Repeat steps 3, 4 and 5 until all of them are optimally clustered.

After using the rotation invariance algorithm, all shape context vectors can be grouped into $k$ clusters. The shape context vectors in the same cluster represent similar shape context with rotation invariance. We then follow the procedure described in section 2 to represent all shapes in the database with the histogram of shape label frequencies. The query image is also replaced with the histogram of shape label frequencies and then the shape matching is performed.

### 4. EXPERIMENTAL RESULTS

We use the ETH-80 image set [21] as our database. The database consists of eight categories, and each category has ten unique objects. Each object is represented by 41 images from viewpoints spaced equally over the upper viewing hemisphere (a distance of $22.5^\circ - 26^\circ$). There are 3280 images totally. For every image, the database provides a high-quality segmentation mask. For a query image, our system will sample 100 points on its contour, determine a shape label to represent it, and then perform image retrieval from our database by utilizing the histogram of shape label frequencies. A retrieved image is considered a match if it belongs to the same category as the query image.

We select 32 rotated images as queries from eight categories, each containing four of the images. Fig. 5 shows some of shortlists of length 5 for the retrieval result in case that the query image is not rotated. We have compared the existing system [22] with the proposed system as shown in Fig. 5. The result of Fig. 5 is reflected as well in the average precision-recall figure (Fig. 6) where low recall indicates few number of the answer set. The precision levels for both methods are similar at low recall levels. When the recall level arises, the precision level for our rotation invariance algorithm is higher than existing method. That is to say, when the number of images in the answer sets is few, the results of two methods are close to each other. But when the number of answer set increases, our rotation invariance algorithm can retrieve more matched images than existing method. The reason is that some images that are different views for the objects in the same category as the query image are similar to the rotations of query image and can only be retrieved by our method.

We further examine the performance of our method by rotating the query images by $90^\circ$. This means that an actual object may be rotated or represented by another view. Fig. 7 shows some of shortlists for rotated query image. It is obvious that our system has better retrieval results. For instance, the answer set of the fourth query in Fig. 7 (a) is not matched with respect to the query image, a rotated cup. On the contrary, the answer set of the fourth query in Fig. 7 (b) is matched with respect to the same query image. Fig. 8 shows the average precision versus recall figures. Significantly, our system has higher precision levels at all recall levels than the existing system.
We also use the MPEG-7 CE [22] database, which is a widely accepted evaluation platform for the rotation invariance matching. The MPEG-7 CE-Shape-1 database consists of 1400 shapes and 70 classes, i.e. each class has 20 shapes. The retrieval accuracy is reported by the so-called Bullseye test: Every shape is matched with all the other shapes in a dataset. We keep 40 shapes with the highest similarity scores and discard the others. Among the 40 retrieved shapes, we count the correct hits, i.e. the number of shapes belonging to the same class as the query shape. There are at most 20 hits for each query. The accuracy of shape retrieval is the ratio of the number of correct hits to the highest possible correct hits (which is $20 \times 1400$ in this experiment).

We conduct sampling uniformly along contours and take $N$ sampling points for each shape. Note that some shapes in the MPEG-7 database have mirror symmetry. To handle
mirrored shapes, we match a query to an original shape and also its mirrored shapes, and keep matches with higher similarity scores. Table 1 shows the retrieval rates and average matching times for various $N$. We observed that the system achieves high speed while maintaining good accuracy.

<table>
<thead>
<tr>
<th>$N$</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrieval rates (%)</td>
<td>71.45</td>
<td>72.23</td>
<td>74.82</td>
<td>76.49</td>
<td>77.54</td>
<td>78.91</td>
</tr>
<tr>
<td>average matching times (ms)</td>
<td>0.68</td>
<td>0.72</td>
<td>0.78</td>
<td>0.84</td>
<td>0.88</td>
<td>0.96</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

In this paper, an efficient image retrieval system is proposed. In this system, we present a rotation invariance algorithm for rotational-invariant shape matching using shape context the algorithm is adopted in the vector quantization process of shape contexts for the image database and groups similar shape contexts of images in the same cluster whether images are rotated or not. Therefore, similar shapes whether they are rotated or not, will have similar shape labels to the query object and can be retrieved by our system. This means our system has better retrieval performance than the existing shape matching method. The experiment on ETH-80 image set confirms that our system has higher precision at all recall levels when the query image is rotated and at most recall levels when the query image remains its original view.

REFERENCES

11. I. Biederman, “Recognition-by-components: A theory of human image understand-


Yen-Tai Lai (賴源泰) received a B.S. from the National Taiwan Ocean University in 1975, an M.S. from National Cheng Kung University, Taiwan, in 1980, both in Electrical Engineering, and a Ph.D. in electrical Engineering and Computer Science from the University of Illinois, Chicago, in 1984. After graduation, he worked at RCA, Solid State Division, Somerville, New Jersey, from 1985 to 1988. He is currently a Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan.
Chi-Chou Kao (高雄洲) received a B.S. in Engineering Science from National Cheng Kung University, Taiwan, in 1990, and M.S. and Ph.D. degrees in Electrical Engineering from National Cheng Kung University, Taiwan, in 1996 and 2002, respectively. Dr. Kao received the Long-Term Distinguished Paper Award, Acer Foundation, in 2003, and appears in the 23rd, 2006 and 24th, 2007 edition of Who’s Who in the World. His research interests include graph algorithms, combinatorial optimization, and circuit design. He is currently an Associate Professor in the Department of Computer Science and Information Engineering, National University of Tainan, Taiwan.

Tzu-Chiang Tai (戴自強) received B.S. and M.S. degrees in the Electrical Engineering from National Cheng Kung University, Taiwan, in 1995 and 1997, respectively. He is currently pursuing the Ph.D. degree in Electrical Engineering, National Cheng Kung University, Taiwan. His research interests include computer-aided design of VLSI circuits, image processing and reconfigurable computing.

Wen-Chun Yeh (葉紋君) received a B.S. in Mechanical Engineering from National Sun Yat-sen University, Taiwan, in 2004, and a M.S. degree in Electrical Engineering from National Cheng Kung University, Taiwan, in 2006. She is currently with the Magma Design Automation, Hsinchu, Taiwan. Her research interests include VLSI design automation, image retrieval, and pattern recognition.