



SHAPE MEASURES FOR CONTENT BASED IMAGE RETRIEVAL: A COMPARISON

BABU M. MEHTRE¹, MOHAN S. KANKANHALLI^{2*} and WING FOON LEE³

¹ New Technologies Group, CMC Centre, Gachibowli, Hyderabad, 500019, India, ² Institute of Systems Science, National University of Singapore, Kent Ridge, Singapore, 119597, Singapore and ³ School of Electrical & Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore, 639798, Singapore

(Received 2 October 1995; accepted 12 September 1996)

Abstract—A great deal of work has been done on the evaluation of information retrieval systems for alphanumeric data. The same thing can not be said about the newly emerging multimedia and image database systems. One of the central concerns in these systems is the automatic characterization of image content and retrieval of images based on similarity of image content. In this paper, we discuss effectiveness of several shape measures for content based similarity retrieval of images. The different shape measures we have implemented include outline based features (chain code based string features, Fourier descriptors, UNL Fourier features), region based features (invariant moments, Zernike moments, pseudo-Zernike moments), and combined features (invariant moments & Fourier descriptors, invariant moments & UNL Fourier features). Given an image, all these shape feature measures (vectors) are computed automatically, and the feature vector can either be used for the retrieval purpose or can be stored in the database for future queries. We have tested all of the above shape features for image retrieval on a database of 500 trademark images. The average retrieval efficiency values computed over a set of fifteen representative queries for all the methods is presented. The output of a sample shape similarity query using all the features is also shown. © 1997 Elsevier Science Ltd

1. INTRODUCTION

There have been many studies in evaluation of information retrieval systems (Harman, 1992; Robertson & Hancock-Beaulieu, 1992; Salton, 1992) for alphanumeric data. Recently, there has been an increasing interest in the area of image and multimedia information systems. An important area of research in these emerging systems is the automatic characterization of image content and retrieval of images based on similarity of image content. In image information systems, a common criterion for data retrieval is “what objects (shapes or images) in the database match or are closely similar to a given shape or image?” This type of retrieval is called shape similarity based retrieval. Given the potentially large size of image databases resulting from the variety of visual sensors currently available, it seems reasonable to address the problem of shape matching with an eye to efficiency (Gary & Mehrotra, 1992).

In this paper, we make an attempt to evaluate (assess the effectiveness) of several shape measures for similarity retrieval of images. Similarity retrieval based on image content forms an important feature for image and multimedia databases (Wu *et al.*, 1995). In fact, there are a couple of commercial systems now available which have implemented variations of shape techniques described in this paper (Flickner *et al.*, 1995; Bach *et al.*, 1996). There are several shape measures (Marshall, 1989; Mehtre & Narasimhalu, 1992) which can be used for recognition and retrieval of images, an overview of which is given in Section 3. An important criterion for shape measures is that the measures be invariant to affine transformations (i.e. rotation, scaling, and translation) of images. This is because, human beings ignore such

* To whom all correspondence should be addressed.

variations in images for recognition and retrieval purposes.

Scassellati *et al.* (1994) have compared shape computation methods with human perceptual judgements for retrieval of images. They have compared the following seven methods of shape computation: algebraic moments; parametric curve distance; parametric curve and first derivative distance; parametric curve, first derivative, and second derivative distance; turning angle; sign of curvature; and modified Hausdorff distance. All these methods (except the algebraic moments) compute shape from the image outline. They have tested the above methods on several query shapes which were all single components (so that they could compute the boundary). They have presented their results as the weighted responses for different methods separately for each query shape.

Our study, though similar in spirit, is different in content and approach. We have studied the following shape measures for their effectiveness of shape similarity retrieval:

- 1. *Outline based features*
 - (a) Chain coded string
 - (b) Fourier descriptors
 - (c) UNL Fourier features
- 2. *Region based features*
 - (a) Invariant moments
 - (b) Zernike moments
 - (c) Pseudo-Zernike moments
- 3. *Combined features*
 - (a) Invariant moments and Fourier descriptors
 - (b) Invariant moments and UNL Fourier features

Our database consists of complex images consisting of several arbitrary shapes. We have tested all the above features on a database of 500 actual trademark images and presented the comparative results. A sample shape similarity retrieval output for a query image as produced by different methods under consideration has also been shown.

The organization of this paper is as follows: we introduce concepts of image retrieval in Section 2, followed by a brief overview of shape recognition techniques in Section 3. Section 4 describes a set of important shape measures along with their corresponding similarity measures. Test results are discussed in Section 5 and conclusions are given in Section 6.

2. IMAGE RETRIEVAL

2.1. The retrieval model

Figure 1 shows a block schematic of a generic image archival and retrieval model. In essence, image matching and retrieval is based on some characteristic features of the image class under

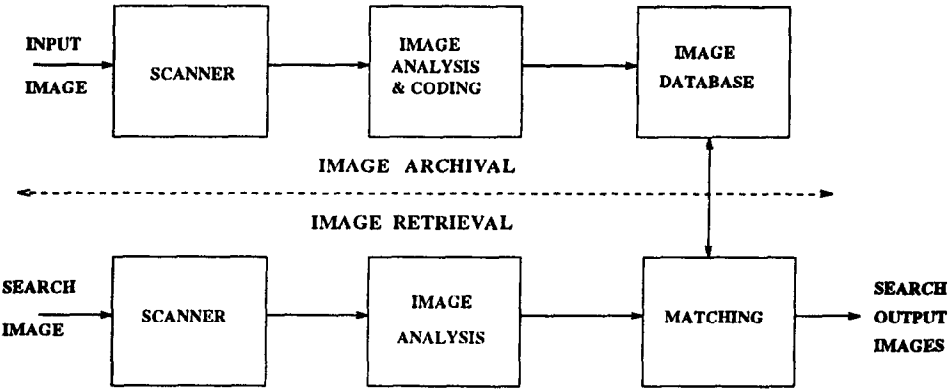


Fig. 1. An image retrieval model.

consideration. The input images are analyzed to extract the features and these features are used to store in the image database, along with the original images. These features could, for example, be shape features, texture features or color features. Whenever an image is submitted for search, it is analyzed and its features are extracted. These extracted features are matched against those in the database. A set of closely matching images are brought out as the result of search output.

An important criterion for testing the efficacy of the search and retrieval is that the output must include all the similar images. The list may have other images as well, but that is not very important. The important thing is that the similar ones should not be missed in the search process. If a similar image is not brought out, it would defeat the purpose of having an automated search. Such a criterion is important in many applications like trademark registration, fingerprint identification etc., where the system brings out the short list and the final decision is taken by the human expert in the loop. Therefore, to evaluate the performance of retrieval we have used a figure of merit called Retrieval Efficiency. The efficiency of retrieval (Mehetre *et al.*, 1995), η_r , for a given shortlist of size T is given by:

$$\eta_r = \begin{cases} \frac{n}{N} & \text{if } N \leq T \\ \frac{n}{T} & \text{if } N > T \end{cases}$$

where n is the number of similar images retrieved and N is the total number of similar images in the database. Note that if $N \leq T$, then η_r reduces to the traditional *recall* measure of information retrieval. And if $N > T$, then η_r computes the *precision* measure of information retrieval.

2.2. Image features and matching

Our problem is as follows: assume that we have a large number of images in the database. Given a query image, we would like to obtain a list of images from the database which are most “similar” in some aspect (here we consider the shape aspect) to the query image. For solving this problem, we need two things—first, a feature which represents the shape information of the image and second, a similarity measure to compute the similarity between features of two images. The similarity measure is a matching function, and gives the degree of match or similarity for a given pair of images (represented by shape measures). The desirable property of a similarity measure is that it should be a metric (that is, it has the properties of symmetry, transitivity and linearity).

Therefore, for the purpose of matching and similarity computation, an image I can be represented by a feature vector \mathbf{f}' which captures different aspects of image content such as color information, texture features, shape information of objects in the image, spatial layout of the objects or any other content which may be of interest for the purpose of querying. Thus the image I can be represented as:

$$\mathbf{f}' = (i_1, i_2, \dots, i_n)$$

where n is the number of content features.

Once the image information is captured by the defined feature vector, the important thing to be determined is the similarity between any two images. The similarity between two images is computed using the feature vectors for any type of content-based similarity retrieval. If we assume that the feature space is uniform, then the similarity measure between a pair of images Q and I having feature vectors \mathbf{f}^Q and \mathbf{f}' can be computed as the Euclidean distance between the two feature vectors. The uniformity assumption of the feature space implies that perceptual distances between points in the space correspond to the Euclidean metric. The similarity measure is therefore:

$$d(Q, I) = \sqrt{\sum_{j=1}^n (\mathbf{f}_j^Q - \mathbf{f}_j')^2}$$

The distance between two identical images is zero, i.e. $d(Q,Q)=0$. Smaller values of distance $d()$ indicate more similarity between images and vice-versa. For similarity retrieval of images, the Euclidean distance $d(Q,I)$, can be computed between the query image and all the database images. The list can then be sorted based on the value of the distance in the ascending order. The output of such a retrieval is known as the ranked similarity output.

3. AN OVERVIEW OF SHAPE RECOGNITION TECHNIQUES

There are many techniques of shape description and recognition. These methods can be categorized as information preserving (IP) or unambiguous, and non-information preserving (NIP) or ambiguous, depending on whether or not it is possible to reconstruct a reasonable approximation of the object from the shape descriptors (Pavlidis, 1980; Pernus *et al.*, 1992). The IP methods (such as Khotanzad & Hong, 1991; Kim & Park, 1990) are mostly used in data compression applications, where there is a need to reconstruct the original image. We shall concentrate on NIP methods which are very useful as approximate indicators of shape, although they do not permit the reconstruction of the original object.

An overview of shape description techniques is given in Fig. 2. These techniques can be broadly categorized into two types, boundary based and region based. Boundary based methods use only the contour or border of the object shape and completely ignore its interior. Hence, these methods are also called external methods. The region based techniques take into account internal details (like holes etc) besides the boundary details. Recognition of a shape by its boundary is the process of comparing and recognizing shapes by analyzing the shapes' boundaries; but the local structural organization is always hard to describe.

Generally speaking, shape recognition has two major parts: shape representation and shape

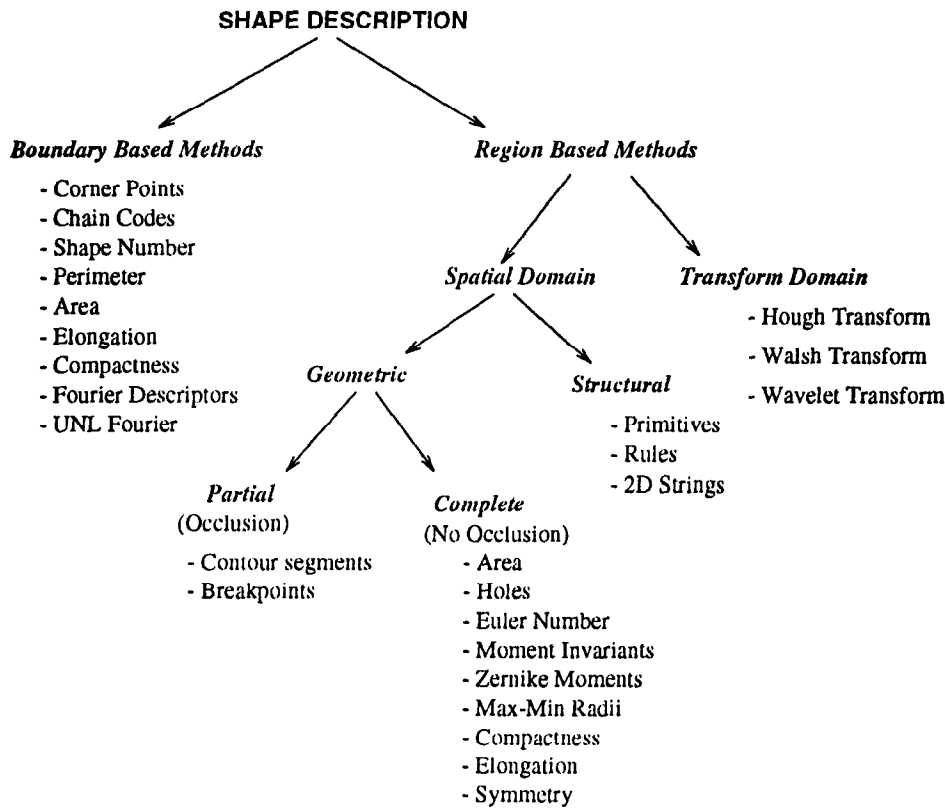


Fig. 2. An overview of shape description techniques.

matching. Shape representation has been widely discussed in recent decades (Blatt & Rubinstein, 1994). Chain coding (Bues & Tu, 1987; Freeman & Davis, 1977), Fourier descriptors (Lin & Hwang, 1987), invariant moments (Hu, 1962; Jiang & Bunke, 1991), autoregressive models (Dubois & Glanz, 1986), polar signatures (Lie & Chen, 1986) and syntactic approaches (Chen & Su, 1986) are all well known methods of shape representation. Among them, Fourier descriptors, invariant moments, autoregressive models and polar signature methods are independent of translation and rotation. As for shape matching, numerous approaches have also been proposed based upon the shape representation methods.

The objective of shape measures (or shape descriptors) is to measure geometric attributes of an object, that can be used for quantifying shapes (Gonzalez & Wintz, 1987; Jian, 1989; Rosenfeld & Kak, 1982), matching shapes, and recognizing objects. Some of the simpler attributes are length, area and radius of shape boundary. The length attribute is also called perimeter of the boundary, and can be computed by just counting the number of pixels on the boundary. The disadvantage of such simple measures is that they are scale and size dependent, whereas the concept of shape is invariant to scale, size and orientation. Measures such as corners, Euler number, roundness (or compactness), symmetry, and shape number are invariant to the factors mentioned. The Euler number (Gonzalez & Wintz, 1987) is defined as difference between the numbers of connected regions and of holes in an object. Roundness (Jian, 1989) is defined as the ratio of the square of perimeter to the area. For circular objects this factor is close to 1 (equal to 1 for circle); for elongated objects it is large.

Recently much attention has been paid to the design of image databases (Chang *et al.*, 1988; Hermes *et al.*, 1994; Rabitti & Savino, 1991; Tegolo, 1994; Ang & Desai, 1993; Chang & Lee, 1991; Grosky & Mehrotra, 1990; Jagadish, 1991; Lee *et al.*, 1989). Due to advances in imaging technologies and retrieval methods many new application areas involving image information retrieval are emerging. In the following section we discuss some important shape features suitable for image retrieval.

4. SHAPE FEATURES FOR IMAGE RETRIEVAL

The shape or form of the object is an attribute that must not change when the original object is submitted to a certain set of affine geometric transformations, or an arbitrary combination of the same. A 2-D shape descriptor should be insensitive to:

- Translation
- Scale changes (uniform in both the X -coordinate and the Y -coordinate)
- Rotations

This criterion for shape descriptors implies that the descriptor is able to perform some normalization for the different appearances in which the object may occur. The actual instantiation of the object must be mappable to an isomorphous prototype of the pattern.

We now briefly describe each of the studied shape feature and the corresponding similarity measure in the following sections. A general source for all these techniques is (Nadler & Smith, 1992).

4.1. Outline based features

4.1.1. Chain coded string. The chain code (also known as Freeman code) can be used for representing the boundary of any shape (Freeman & Davis, 1977). The boundary can be traced in either a clockwise or a counter-clockwise manner and eight codes for every pixel are assigned according to the direction of the next pixel with respect to the current pixel. The eight direction codes are shown in Fig. 3(a). Let $C=(a_1a_2...a_n)$ be the chain code of a contour. The chain codes have the modulo-8 property, i.e. $a_8=a_0$ [see Fig. 3(a)]. Since the ordinary chain code is not invariant to scale and rotation, we used some modified chain codes for the purpose of shape description. We can then apply similarity measures on these modified chain code representa-

tions. The reduced chain code representation (*RC*) of that contour is $RC=(b_1b_2... b_m)$ where $m=n/2$, which has the modulo-4 property, i.e. $b_4=b_0$. The reduced difference chain code is obtained by the difference of adjacent codes, that is $RDC=b_i-b_{i+1}$ [see Fig. 3(b)]. It should be noted that *RDC* is better suited for contour representation for the purpose of shape description and matching than the chain code itself. This is because it is invariant to rotation and scaling.

Since the chain code is not unique (as it varies with rotation and scaling), another code called the derivative chain code was developed by Bribiesca and Guzman (Bribiesca & Guzman, 1978) which is unique. Instead of using the orientations of chain links, they use what they call “the derivative of the chain code” and is formulated as follows:

1. Two successive links in the same direction are encoded by 2.
2. A convex corner is encoded by 1.
3. A concave corner is encoded by 3.

See Fig. 3(c) for an illustration. The number of digits in the code is called the “order”. Shape codes must be of even order. To obtain a unique shape code, one arbitrary encoding is selected and all cyclic permutations are generated from it. The unique code is the one among these permutations such that the ternary number interpretation of this code is the smallest.

Similarity measure

An image consists of objects which have contours (boundaries). The image contours are represented by chain codes. These chain codes are considered to be the strings describing the shapes. For matching a pair of image outlines (shape contours), their string representations are compared. This can be done using the string distance measures. There are the following three types of distance measures: weighted Levenstein distance (*WLD*), nonlinear elastic matching

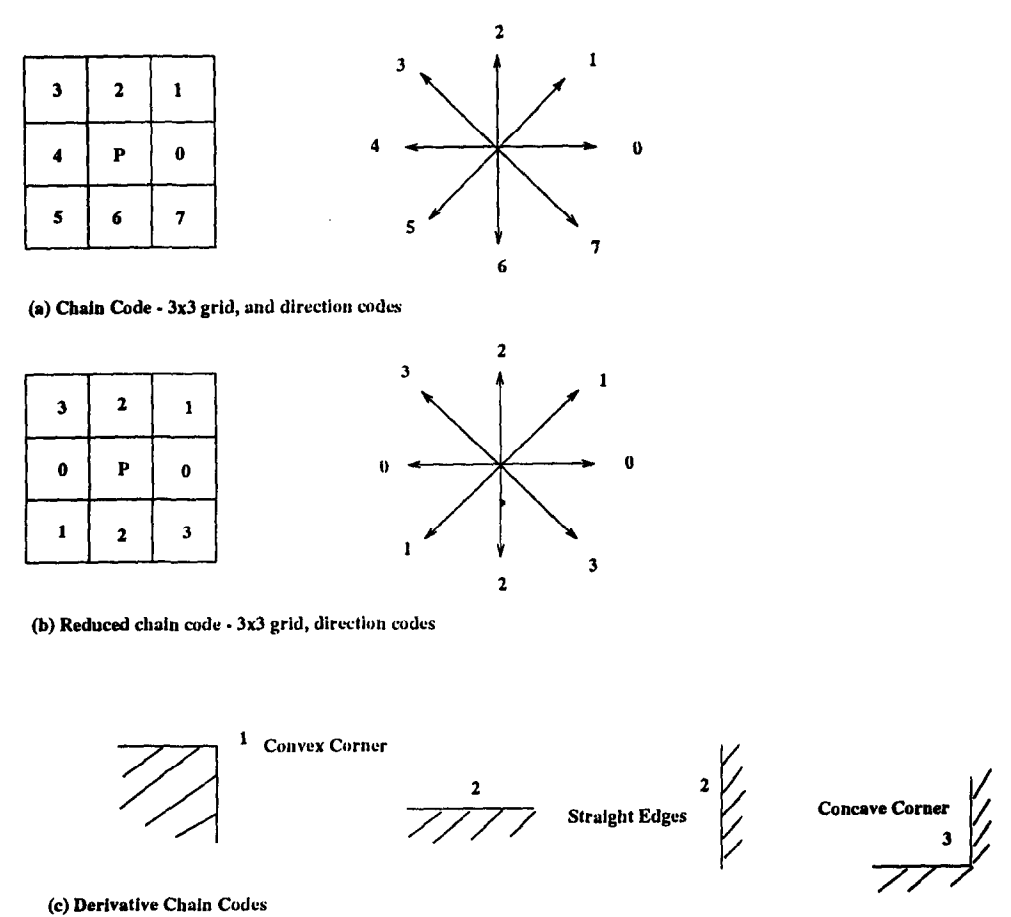


Fig. 3. Illustration of chain code and its variations.

(NEM), and extended distance (ED) (Cortelazzo *et al.*, 1994). We explain these measures below.

Consider two strings A and B :

$$A = \{a_1, a_2, \dots, a_n\} \quad B = \{b_1, b_2, \dots, b_m\}$$

If s_k , i_k and d_k substitutions, insertions and deletions are required to rewrite string A into B and p , q & r are the relative costs of substitutions, insertions and deletions respectively, then the WLD distance between strings A and B is defined as:

$$D_{\text{WLD}} = \min_k \{s_k p + i_k q + d_k r\}$$

The NEM distance is defined as:

$$D_{\text{NEM}} = l(m, n)$$

where

$$l(i, j) = \min(l(i-1, j-1) + w_s, l(i, j-1) + w_i, l(i-1, j) + w_d)$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$ with

$$w_s = \begin{cases} 0 & \text{if } a_i = b_j \\ p & \text{otherwise} \end{cases} \quad w_i = q, \quad w_d = r.$$

and boundary conditions

$$l(i, 0) = i r \quad i = 0, 1, 2, \dots, m$$

$$l(0, j) = j q \quad j = 0, 1, 2, \dots, n$$

The ED distance can be similarly defined as:

$$D_{\text{ED}} = l(m, n)$$

with

$$w_s = \begin{cases} 0 & \text{if } a_i = b_j \\ \alpha p & \text{if } ((a_{i-1} = a_i = b_{j-1}) \text{ and } (a_{i+1} = b_j = b_{j+1})) \text{ or} \\ & \text{if } ((a_{i-1} = b_j = b_{j-1}) \text{ and } (a_{i+1} = a_i = b_{j+1})) \\ p & \text{otherwise} \end{cases}$$

$$w_i = \begin{cases} \beta q & \text{if } a_i = b_{j-1} = b_j \\ p & \text{otherwise} \end{cases}$$

$$w_d = \begin{cases} \gamma r & \text{if } a_{i-1} = a_i = b_j \\ r & \text{otherwise} \end{cases}$$

where α , β , γ are real numbers in the closed interval $[0, 1]$. The details can be found in (Cortelazzo *et al.*, 1994).

4.1.2. Fourier descriptors. Fourier descriptors are complex coefficients of the Fourier series expansion of waveforms (Jian, 1989; Persoon & Fu, 1977). From the boundary trace of a shape, a pair of one-dimensional waveforms $x(t)$, $y(t)$ can be generated. If N samples of a closed boundary are taken, then

$$u(n) \triangleq x(n) + jy(n), \quad n = 0, 1, \dots, N-1 \quad (1)$$

will be periodic with a period N . Its discrete Fourier Transform (DFT) representation is

$$f(k) \triangleq \sum_{n=0}^{N-1} u(n) \exp\left(\frac{-j2\pi kn}{N}\right), \quad 0 \leq k \leq N-1 \quad (2)$$

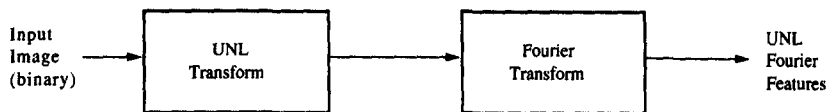


Fig. 4. Illustration of UNL Fourier features computation.

$$u(n) \triangleq \frac{1}{N} \sum_{k=0}^{N-1} f(k) \exp \left(\frac{j2\pi kn}{N} \right), 0 \leq n \leq N-1 \quad (3)$$

The complex coefficients $f(k)$ are called the Fourier descriptors (FDs) of the boundary. The FDs are invariant to the starting point of sampling, rotation, scaling as well as reflection.

Similarity measure

Suppose $F^I = (f_1^I, f_2^I, \dots, f_n^I)$ and $F^Q = (f_1^Q, f_2^Q, \dots, f_n^Q)$ are Fourier descriptors for a database image and a query image respectively. Then the similarity measure is computed as the Euclidean distance between them as follows:

$$d(F^I, F^Q) = \sqrt{\sum_{j=1}^n (f_j^Q - f_j^I)^2}$$

where n is the number of Fourier descriptors.

4.1.3. UNL Fourier features. UNL¹ Fourier features are an extension and improvement over the Fourier descriptors in that they handle open curves, lines etc. The UNL Fourier features (Rauber & Steiger-Garção, 1992; Rauber, 1994) are computed in two stages. This is illustrated in Fig. 4. In the first step, the input image (consisting of binary curve patterns) is transformed from the Cartesian coordinate system to the polar coordinate system by the UNL Transform. For this the analytic curve equations are first estimated, then the analytical UNL transform is performed and finally the transformed curves are instantiated in the polar coordinate system. This results in the normalization in terms of the translation and scale changes in the original pattern. This eliminates translation and scale that might have occurred in a given instance of an image.

Rotations of the original pattern result in the cyclic shift of the UNL transformed pattern. This is eliminated in the second processing step, by the 2-D Fourier transform. The Fourier transform takes the image (trace) of the UNL Transformed curve pattern as input. The spectrum of the Fourier Transform is insensitive to any cyclic shifts of its input pattern, hence eliminating the rotation of the original pattern. These characteristics of the UNL Fourier Features (UFF) make it invariant to translation, scale changes and rotation of the original pattern.

Let $\hat{\Omega}(t)$ be a discrete object composed of n pixels $z_i = (x_i, y_i)$. Let $\hat{O} = (\hat{O}_x, \hat{O}_y)$ be the centroid of the object and let \hat{M} be the maximum Euclidean distance from the centroid $\hat{\Omega}$ to all pixels z_i .

The normalizing mapping of object $U(\hat{\Omega}(t))$ —the line segments $z_{ij}(t)$ between two neighboring pixels $z_i = (x_i, y_i)$ and $z_j = (x_j, y_j)$ that compose the object, $U(z - ij(t))$ from the Cartesian to the polar coordinate system, using the following formula is the UNL transform of the discrete object:

$$U: ((0,1) \rightarrow R^2) \rightarrow ((0,1) \rightarrow R^2) \quad \forall \dots n \quad t \in [0,1]$$

$$U(z_{ij}(t)) = \zeta_{ij} = (E_{ij}(t), \theta_{ij}(t)) = \left(\frac{\|z_i + t(z_j - z_i) - \hat{O}\|}{\hat{M}}, \arctan \left(\frac{y_i + t(y_j - y_i) - \hat{O}_y}{x_i + t(x_j - x_i) - \hat{O}_x} \right) \right)$$

Let $i(x,y)$ be a 2-D image that represents a discrete object $\hat{\Omega}(t)$ and let $f(R,\theta)$ be the two dimensional image that represents the UNL transform of $\hat{\Omega}(t)$. The normalized discrete Fourier spectrum

¹Universidade Nova de Lisboa

$$UFF(u', v') \equiv \frac{\|\mathcal{F}\{f(R, \theta)\}\|}{F(0,0)} = \frac{\|F(u, v)\|}{F(0,0)}$$

of the image $f(R, \theta)$, ignoring $F(0,0)$ and the values that are duplicated by conjugate symmetry are the discrete UNL Fourier features of the object $\hat{f}(t)$.

Similarity measure

Suppose $U^I = (u_1^I, u_2^I, \dots, u_n^I)$ and $U^Q = (u_1^Q, u_2^Q, \dots, u_n^Q)$ are UNL Fourier features for database image and a query image. Then the similarity measure is computed as the Euclidean distance between them as follows:

$$d(U^I, U^Q) = \sqrt{\sum_{j=1}^n (u_j^Q - u_j^I)^2}$$

where n is the number of UNL Fourier features.

4.2. Region based features

4.2.1. Invariant moments. Invariant moments (Hu, 1962) are a very popular shape measure. The 2-D moments of order (p, q) of some function $f(x, y)$ are defined by

$$m_{pq} = \iint x^p y^q f(x, y) dx dy \quad (4)$$

the first moment μ_{00} is denoted by m . Setting $\bar{x} = \mu_{10}/m$, $\bar{y} = \mu_{01}/m$, the central moments of order (p, q) are defined as

$$\mu_{pq} = \iint (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (5)$$

For a digital image, the integrals are replaced by summations to get

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (6)$$

The normalized central moments (Hall, 1979), denoted by η_{pq} , are defined as $\eta_{pq} = \mu_{pq} / (\mu_{00})^\gamma$ where

$$\gamma = \frac{1}{2}(p+q)+1 \text{ for } p+q=2, 3, \dots$$

From the second and third order moments, a set of seven invariant moments can be derived. They are given by:

$$\begin{aligned} m_1 &= \eta_{20} + \eta_{02} \\ m_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ m_3 &= (\eta_{30} - \eta_{02})^2 + (3\eta_{21} - \eta_{03})^2 \\ m_4 &= (\eta_{30} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \\ m_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ m_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ m_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] \\ &\quad + [(3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]] \end{aligned}$$

Thus, the shape feature vector using moment invariants is $f_m = (m_1, m_2, \dots, m_7)$. It may be noted that the invariant moments as shape measure have the desirable property of being invariant under variations of the image content such as shift, scaling, and rotation (referred to as affine transformations). They have been given considerable attention in the literature and reports of their satisfactory experimental results [e.g. aircraft recognition (Dudani, 1977), inaccuracies

(Lin & Wang, 1994; Reiss, 1991), extensions (Maitra, 1979), variations (Teague, 1980; Chen, 1993; Flusser & Suk, 1993; Mustafa & Psaltis, 1984; Smith & Wright, 1971)] have made it appropriate to settle the question how good are moment invariants, in practical applications.

Similarity measure

In case of the moment invariant values, often some moment values are quite small (even zero in many cases). Hence, the similarity measure for moment invariants should take care of this type of variation. And among the moment invariant themselves, their magnitude varies a lot. So, they need to be normalized for comparison. This normalization can be done as follows: the moment invariant values for the entire collection of images is scanned and the limits of each moment invariant is computed. These limiting values are used to normalize the moment invariants between 0 and 1.

The similarity distance (d_m) between two such feature vectors M^Q and M^I for a pair of images Q and I is computed as the Euclidean distance as follows:

$$d_m(M^Q, M^I) = \sqrt{\sum_{i=1}^7 (m_i^Q - m_i^I)^2}$$

The value of d_m is zero (or small) for identical (or similar) images and high for other images.

4.2.2. Zernike moments. Khotanzad and Hong (1990) and Khotanzad & Lu (1990) proposed Zernike moments for image recognition. Zernike moments are derived from Zernike polynomials which form a complete orthogonal set over the interior of the unit circle, i.e. $x^2 + y^2 = 1$. Let the set of these polynomials be denoted by $V_{nm}(x, y)$. The form of these polynomials is:

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta) \quad (7)$$

where

- n Positive integer or zero.
- m Positive and negative integers subject to constraints $n - |m|$ even, $|m| \leq n$.
- ρ Length of vector from origin to (x, y) pixel.
- θ Angle between vector ρ and x axis in counterclockwise direction.
- $R_{nm}(\rho)$ Radial polynomial defined as

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s \right)! \left(\frac{n-|m|}{2} - s \right)!} \rho^{n-2s} \quad (8)$$

Note that $R_{n,-m}(\rho) = R_{nm}(\rho)$. The Zernike moment of order n with repetition m for a continuous image function $f(x, y)$ that vanishes outside the unit circle is

$$A_{nm} = \frac{n+1}{\pi} \int_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(\rho, \theta) dx dy \quad (9)$$

For a digital image, the integrals are replaced by summations to get

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(\rho, \theta), x^2 + y^2 \leq 1. \quad (10)$$

To compute the Zernike moments of a given image, the center of the image is taken as the origin and pixel coordinates are mapped to the range of unit circle, i.e. $x^2 + y^2 \leq 1$. Those pixels falling outside the unit circle are not used in computation. Also, note that $A_{nm}^* = A_{n,-m}$.

4.2.3. Image normalization for Zernike moments. The Zernike (and pseudo-Zernike) moments are invariant only to the rotation, and are not invariant to scaling and translation. So, for effective usage of Zernike moments in shape recognition, the input images need to be normalized for scale and translation. Another important factor to be noted is that Zernike moments take into account only the image within the unit circle. This requires that all the images need to be scaled to a unit circle during the normalization process.

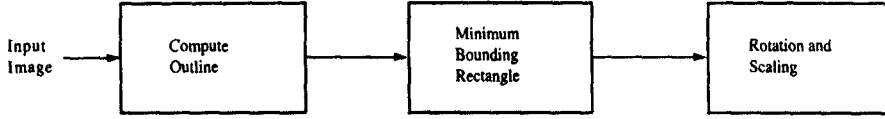


Fig. 5.

Translation invariance can be achieved by shifting the origin to the center of every given image. For scale invariance, the input image is scaled to a square of size M pixel, and the pixel distances are divided by $\sqrt{(2M^2)}$, so as to make it be within unit circle range. Since this scaling is into a unit square, there is loss of information of the aspect ratio. A circle and an ellipse will not be distinguishable after this scaling. So, to keep the information we need to compute the aspect ratio before scaling.

An image can be submitted to the system in any orientation, blindly scaling in the above manner will result in different shape and hence different Zernike moment values. To avoid this problem, and make it invariant to the orientation during submission, it is necessary to compute the minimum bounding rectangle of the image. Thus, the overall scheme of this normalization is illustrated in Fig. 5.

Similarity measure

Let $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$ be Zernike moment vectors for a pair of images A and B , where n is size of the vector. Let $R = \{r_1, r_2, \dots, r_n\}$ be a ratio vector of A and B where $r_i = a_i/b_i$. We compute the variance of R as the similarity distance (Sim_{dist}) between two images as follows:

$$Sim_{dist} = (1/n) \sum_1^n (r_i - R_{avg})^2, \text{ where } R_{avg} = (1/n) \sum_1^n r_i$$

The Sim_{dist} is zero (or nearly zero) for identical (or similar) images and higher for non-similar images.

4.2.4. Pseudo-Zernike moments (PM). Zernike polynomials are polynomials in x and y . A related orthogonal set of polynomials in x , y and r which has properties analogous to those of Zernike polynomials. This set of polynomials, which are referred to as pseudo-Zernike polynomials (Teh & Chin, 1988; Raveendran & Omatu, 1993), differs from that of Zernike in that the radial polynomials are defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|} (-1)^s \frac{(2n+1-s)!}{s!(n-|m|-s)!(n+|m|+1-s)!} \rho^{n-s} \quad (11)$$

where $n=0, 1, \dots, \infty$, and m takes only positive and negative integer values subject to $m \leq n$ only. This set of pseudo-Zernike polynomials contains $(n+1)^2$ linearly independent polynomials of degree $\leq n$, whereas the set of Zernike polynomials contains only $(n+1)(n+2)/2$ linearly independent polynomials of degree $\leq n$.

Similarity measure

The similarity measure for pseudo-Zernike Moments is the same as used in Zernike Moments.

4.3. Combined features

4.3.1. Moment invariant & Fourier descriptors. The shape features using moment invariants and Fourier descriptors are described separately in Sections 4.2.1 and 4.1.2 respectively. The combined shape feature measure for moment invariants and Fourier Descriptors is given by

$$f_c = (f_m, f_f), \text{ where } f_m = (m_1, m_2, \dots, m_7) \text{ and } f_f = (f_1, f_2, \dots, f_n)$$

Similarity measure

$$D_c(Q, I) = \omega_1 d_m + \omega_2 d_f \text{ where } d_m \text{ is given by}$$

$$d_m(M^Q, M^I) = \sqrt{\sum_{i=1}^7 (m_i^Q - m_i^I)^2}$$

and d_f is given by

$$d_f(F^Q, F^I) = \sqrt{\sum_{j=1}^n (f_j^Q - f_j^I)^2}$$

4.3.2. Moment invariant & UNL Fourier features. The combined shape feature measure for moment invariants and UNL Fourier features is given by

$$f_c = (f_m, u_f), \text{ where } f_m = (m_1, m_2, \dots, m_7) \text{ and } u_f = (u_1, u_2, \dots, u_n).$$

Similarity measure

$$D_c(Q, I) = \omega_1 d_m + \omega_2 d_u$$

where d_m is given by

$$d_m(M^Q, M^I) = \sqrt{\sum_{i=1}^7 (m_i^Q - m_i^I)^2}$$

and d_u is given by

$$d_u(U^Q, U^I) = \sqrt{\sum_{j=1}^n (u_j^Q - u_j^I)^2}$$

5. TEST RESULTS AND DISCUSSIONS

We have implemented all the above shape measures in C, on a UNIX workstation (SparcStation 10), and tested on a database of over 500 trademark images. We chose to use trademark images since shape features are important for content based retrieval of trademark images (Wu *et al.*, 1994). Trademarks images are quite complex and are highly heterogeneous. From a trademark registration point of view, when a new trademark is submitted for registration, it is necessary to find out whether a submitted (query) trademark is confusingly similar to any of the existing registered trademarks in the database. This is like issuing a similarity retrieval query for images. This is important because 60% of all the trademarks are (non-alphanumeric) images, also called device marks or figurative elements of marks.

We briefly discuss about the image segmentation module for the purpose of comparison of shape measures. Since trademark images are quite complex and usually have more than one component, the diversity and complexity does not permit automatic segmentation. In fact, often even the foreground and background of images is also not known in advance. The image segmentation process partitions an image into its different components. The scheme comprises a suite of segmentation techniques which may be selected by the system automatically or chosen interactively by the user depending upon the level of complexity of the input image. The images may be two-tone (binary), half-tone (grey), or full color images. The image may contain unlimited variety of patterns including text, graphics, texture etc. The system detects the type of input image and automatically selects an appropriate segmentation method. The results of such partial segmentation are presented to the user. The user then specifies important components of the image, and the features of these specified components are computed. Figure 6 shows the segmentation module of our trademark system (Wu *et al.*, 1994), wherein a trademark image consisting of two components is displayed. Features for the two components are computed and

can be used either for storing in the database or for the query purpose.

To compute the effectiveness of the similarity retrieval of different shape measures, a test was designed as follows: 15 query images which represent the population well were selected. The query images used are shown in Fig. 7. For each query image, a list of similar images present in the 500-images database was first manually found. The above described shape features of the query image were then compared to the corresponding shape features of the images in the database to obtain a short list of similar images. Then the efficiency of retrieval, η_T was calculated. This was computed over several sizes of short lists—5, 10, 15 and 20. Table 1 shows the average retrieval efficiency, η_T computed over 15 queries.

A sample shape similarity retrieval output for a query trademark image is shown in Figs 8, 9 and 10. Here the query image has a star component, and all the database images that have a star in them will be considered as similar images regardless of scale, rotation, or boundary

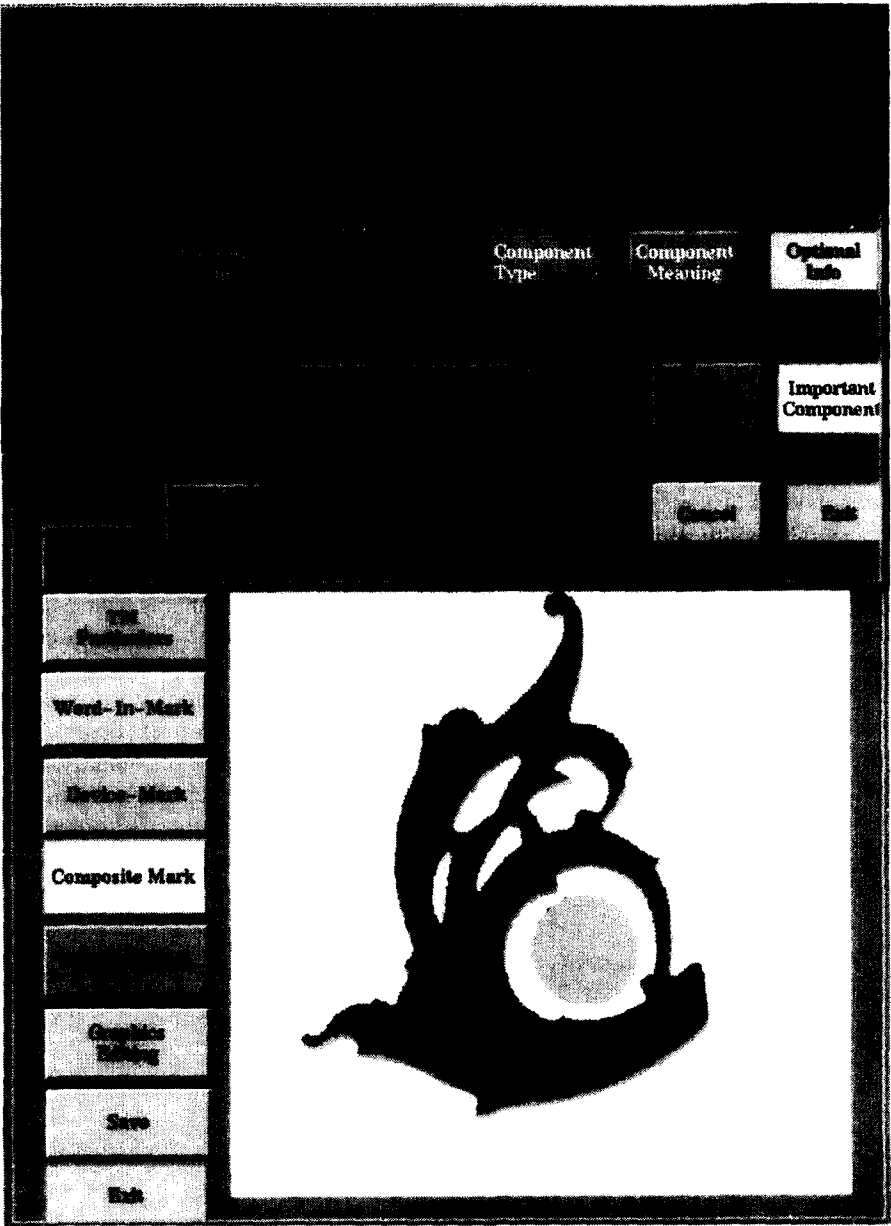


Fig. 6. Segmentation module of STAR, System for Trademark Archival and Retrieval. Two image components are displayed after segmentation.

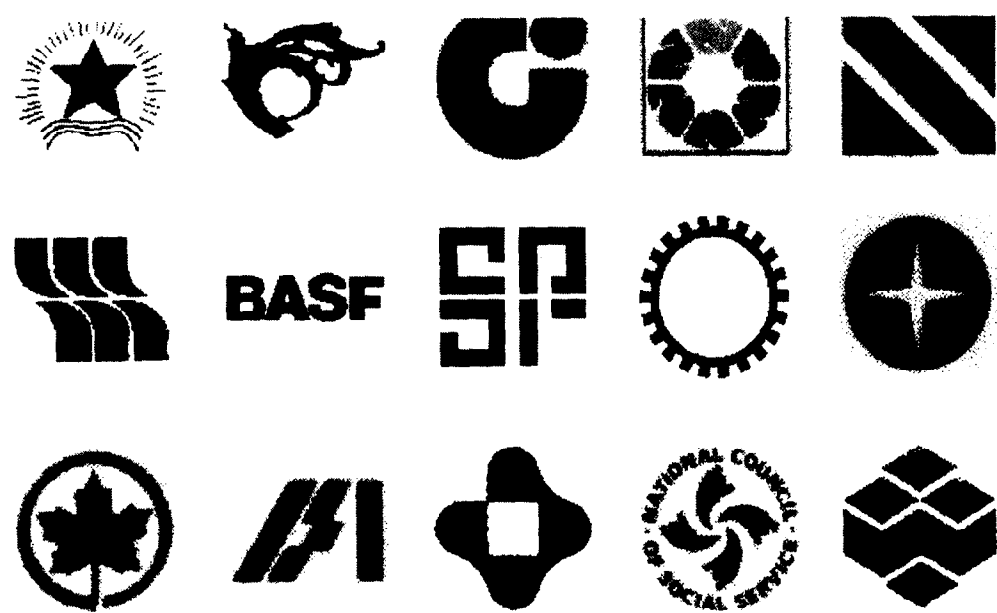


Fig. 7. The set of query images used in our test.

conditions. Figure 8 shows the similarity retrieval output using outline based features, i.e. Reduced chain code, Fourier descriptors, and UNL Fourier features. Figure 9 shows results of shape similarity retrieval using region based features, i.e. Moment invariants, Zernike moments, and pseudo-Zernike Moments. Figure 10 shows results of shape similarity retrieval using combined features, i.e. Moment invariants and Fourier descriptors, and moment invariants & UNL Fourier feature.

It can be seen from these figures that the retrieval output using combined features is better than the individual features. In particular, results of MI and UNL feature are better than MI and FD features. Though there is one false hit (FH) in the case of MI and UNL and none in the MI and FD case, but this is only for the query shown here. Table 1 shows the average values of retrieval efficiencies for all the features computed over 15 queries. From this table MI and UNL feature provides the best retrieval efficiency value. On the other hand, the chain code based string feature has the lowest values of retrieval efficiency. From these experiments, we feel that since the strings comprising of chain-codes of shapes carry more semantic information (in terms of geometry and topology) than say strings in other applications like free-text databases, simple syntactic string distances do not adequately capture the semantics of shape similarity. Also, such a method is not robust with respect to noise introduced by the scanning and digitization processes. Hence chain-code based string-matching methods do not perform as well as the shape similarity techniques based on Fourier descriptors and moment invariants. As far as the moment based methods go, the traditional MI method gives the best performance. This may be due to the

Table 1. Average retrieval efficiency values for different methods. *T* is the short list size of retrieved images

Methods	T=5	T=10	T=15	T=20
Reduced chain code	55.1%	47.6%	50.0%	60.6%
Fourier descriptors (FD)	72.2%	76.9%	75.9%	74.9%
UNL features	81.3%	79.9%	83.7%	89.3%
Moment invariants (MI)	84.7%	86.3%	86.8%	87.7%
Zernike moments	66.9%	66.5%	70.4%	78.2%
Pseudo-Zernike moments	66.9%	66.5%	70.4%	78.2%
MI and FD	93.8%	87.3%	87.1%	89.6%
MI and UNL	93.3%	89.2%	89.3%	91.1%

fact that the normalization process for Zernike and pseudo-Zernike moments may cause some loss of information which is not adequately compensated for by the additional aspect ratio measure. For the outline based methods, UNL has a better performance than FD perhaps due to the increased accuracy of the outline due to the use of analytical curve equations in order to

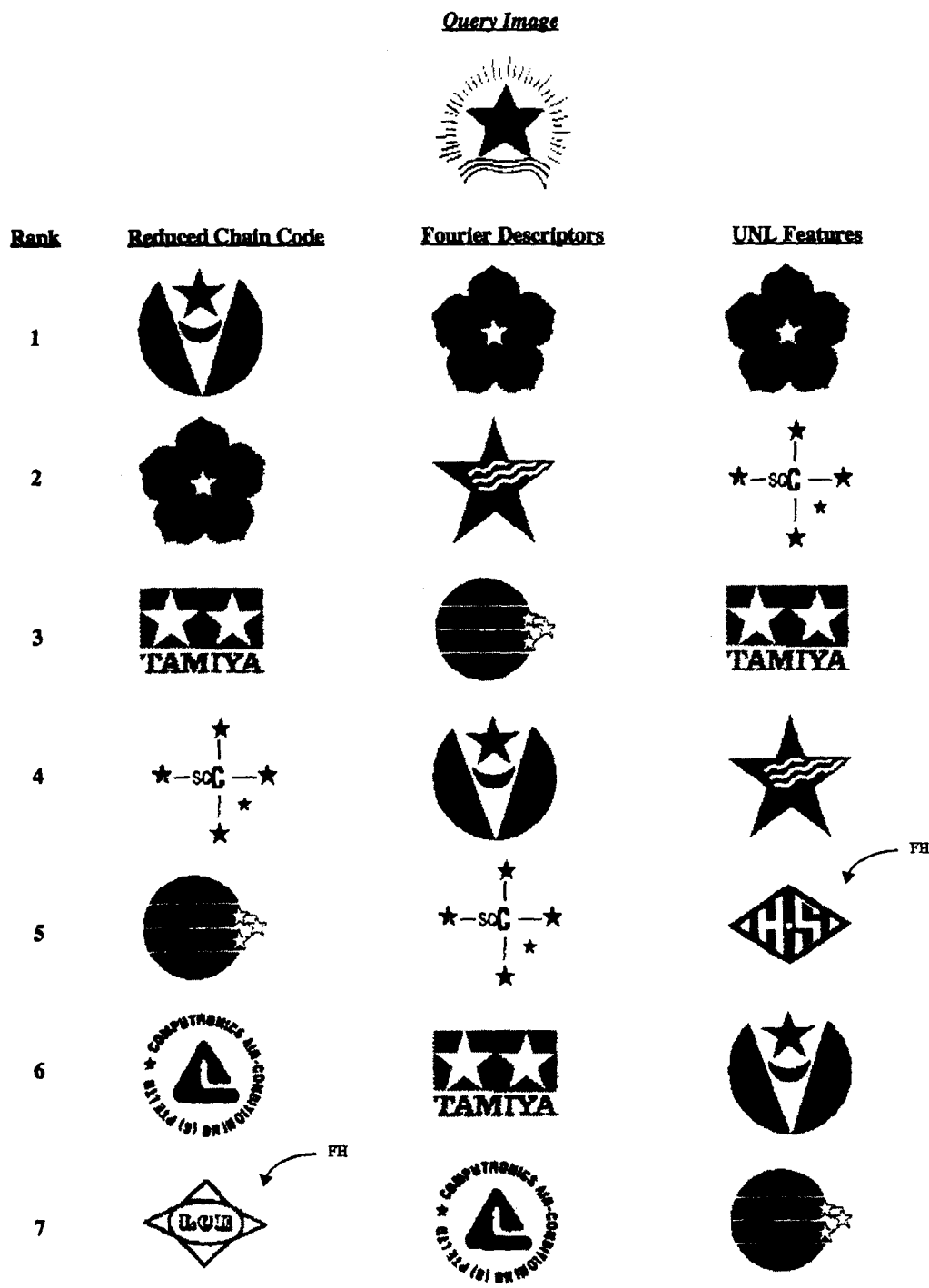


Fig. 8. Results of similarity retrieval using outline based features, for the *star* query image. FH indicates a false hit.

describe it. Finally, our experiments show that a combination of an outline method with a region based method gives best results. This could be because the human perceptual mechanism uses both these aspects of shape in order to compute similarity. In the future, we hope to continue more such experiments on larger databases. We also believe that a deeper study of the human perceptual mechanism would aid in developing more accurate mathematical shape description techniques.

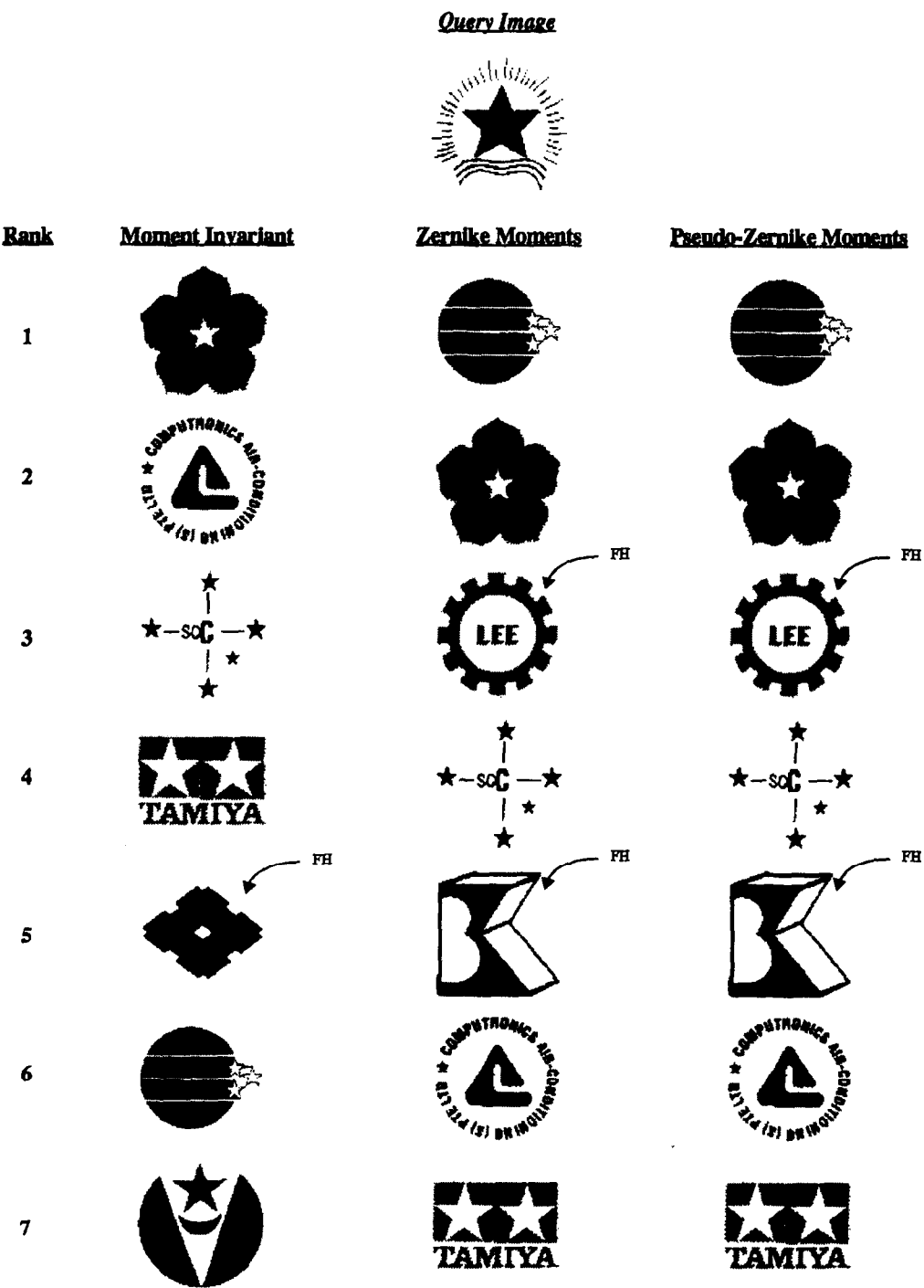


Fig. 9. Results of similarity retrieval using region based features, for the *star* query image. FH indicates a false hit.

6. CONCLUSIONS

We have discussed a set of eight shape measures for the purpose of content based image retrieval. These are outline based features: chain coded strings, Fourier descriptors, UNL Fourier features; region based features: invariant moments, Zernike moments, and pseudo-Zernike moments; and combined features: moment invariants and Fourier descriptors, and moment

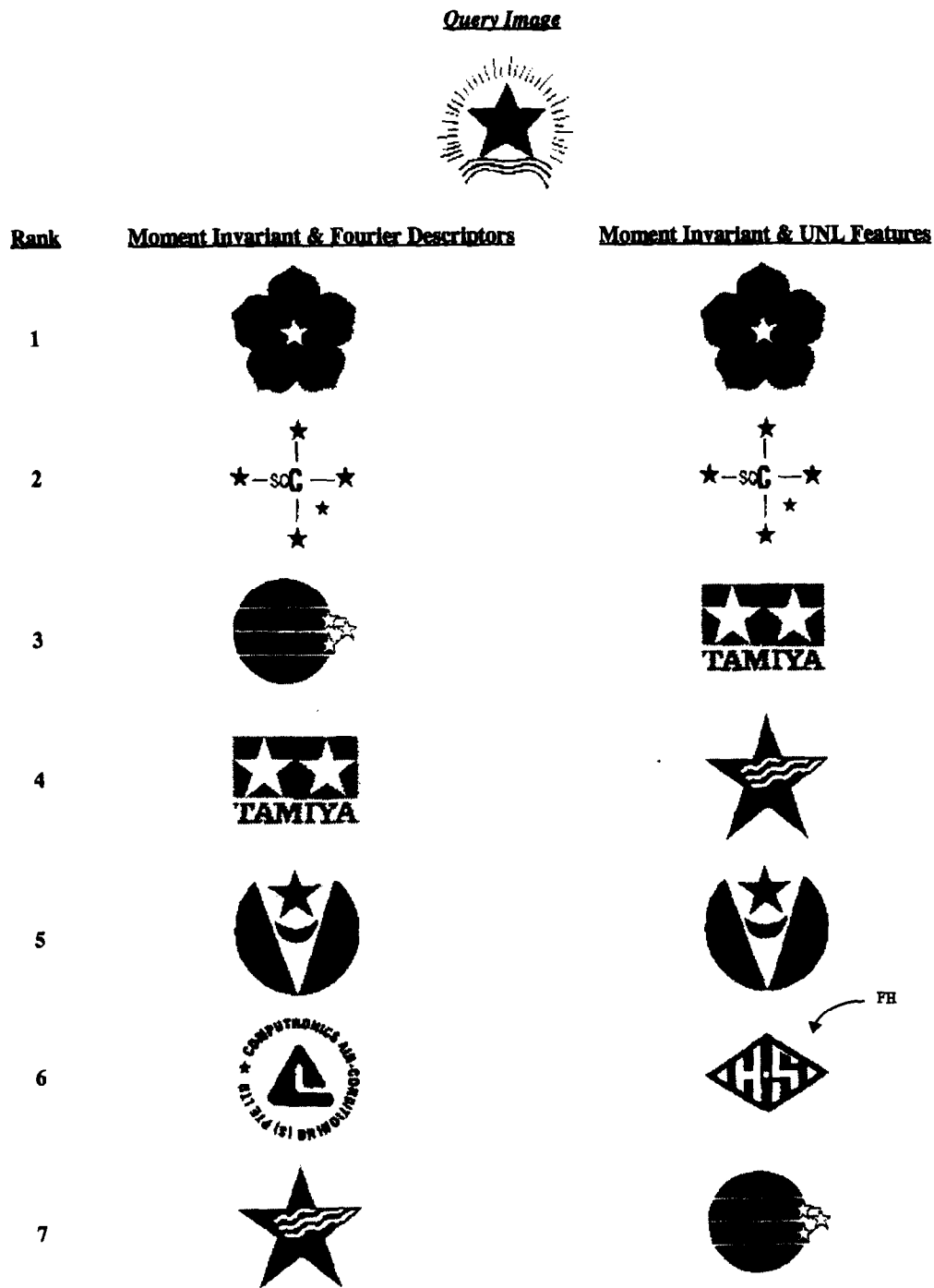


Fig. 10. Results of similarity retrieval using two different combinations of shape features. FH indicates a false hit.

invariants and UNL Fourier features. We have defined the similarity (matching) measures for all these features. The scheme has been implemented in C on a UNIX workstation, and tested on a database of 500 trademark images.

We have tested the effectiveness of these shape measures for the purpose of content based shape similarity retrieval of images, and presented the retrieval efficiency (the figure of merit for image retrieval) values for all the methods. Our test results show that combined features are better than the individual features for effective image retrieval. In particular, the combined feature consisting of moment invariants and UNL Fourier feature gives the best results (highest value of retrieval efficiency).

On the other hand, chain code based string feature gave the lowest value for retrieval efficiency.

Acknowledgements—Mohan S. Kankanhalli's work has been supported by the "Real World Computing Partnership" of Japan. We also thank the reviewers for the encouragement and their perceptive comments which have improved the quality of the paper.

REFERENCES

- Ang, Y. H., & Desai, A. D. (1993). Image information retrieval systems. In Chen *et al.* (Eds.), *Handbook of pattern recognition and computer vision* (pp. 719–740). Cleveland, OH: World Scientific Publishers.
- Bach, J.R. *et al.* (1996). *The Virage image search engine: an open framework for image management*. Paper presented at the SPIE Conference on Storage and Retrieval for Still Image and Video Databases IV.
- Blatt, N., & Rubinstein, J. (1994). The canonical coordinates method for pattern recognition—II. Isomorphism with affine transformations. *Pattern Recognition*, 27(1), 99–107.
- Bribiesca, E. & Guzman, A. (1978). *Shape description and shape similarity for two dimensional regions*. Paper presented at the 4th International Conference on Pattern Recognition, Kyoto, Japan.
- Bues, H. L., & Tu, S. H. (1987). An improved corner detection algorithm based on chain coded plane curve. *Pattern Recognition*, 20(3), 291–296.
- Chang, C. C., & Lee, S. Y. (1991). Retrieval of similar pictures on pictorial databases. *Pattern Recognition*, 24(7), 675–680.
- Chang, S. K., Yan, C. W., Dimitrof, D., & Arnd, T. (1988). An intelligent image database system. *IEEE Transactions on Software Engineering*, 14, 681–688.
- Chen, C. C. (1993). Improved moment invariants for shape discrimination. *Pattern Recognition*, 26(5), 683–686.
- Chen, H. H., & Su, J. S. (1986). *A syntactic approach to shape recognition*. Paper presented at the International Computer Symposium, Tainan, Taiwan.
- Cortelazzo, G., Mian, G. A., Vezzi, G., & Zamperoni, P. (1994). Trademark shapes description by string-matching techniques. *Pattern Recognition*, 27(8), 1005–1018.
- Dubois, S. R., & Glanz, F. H. (1986). An autoregressive model approach to two-dimensional shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 55–66.
- Dudani, S. (1977). Aircraft identification by moment invariants. *IEEE Transactions on Computers*, 26, 39–45.
- Flickner, M. *et al.*, (1995). Query by image and video content: the QBIC system. *IEEE Computer*, 28(9), 23–32.
- Flusser, J., & Suk, T. (1993). Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1), 167–174.
- Freeman, H., & Davis, L. S. (1977). A corner finding algorithm for chain coded curves. *IEEE Transactions on Computers*, 26, 297–303.
- Gary, J. E., & Mehrotra, R. (1992). *Shape similarity-based retrieval in image database systems*. Paper presented at the SPIE Conference on Image Storage and Retrieval Systems No. 1662.
- Gonzalez, R. C., & Wintz, P. (1987). *Digital image processing*, (2nd Edition, Chapter 8). Reading, MA: Addison-Wesley Publishing Co..
- Grosky, W. I., & Mehrotra, R. (1990). Index based object recognition in pictorial data management. *Computer Vision, Graphics, and Image Processing*, 52, 416–436.
- Hall, E. L. (1979). *Computer Image Processing and Recognition*. New York: Academic Press.
- Harman, D. (1992). Evaluation issues in information retrieval. *Information Processing & Management*, 28(4), 439–440.
- Hermes, T., Klauck, C., Kreyb, J., & Zhang, J. (1994). *Image retrieval for information systems*. Paper presented at the SPIE Conference on Storage and Retrieval for Image and Video Databases III.
- Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8, 179–187.
- Jagadish, H. V. (1991). *A retrieval technique for similar shapes*. Paper presented at the ACM International Conference on Management of Data-SIGMOD.
- Jian, A. K. (1989). *Fundamentals of digital image processing*. Chapter 9 on Image Analysis and Computer Vision, Prentice Hall International.
- Jiang, X. Y., & Bunke, H. (1991). Simple and fast computation of moments. *Pattern Recognition*, 24(8), 801–806.
- Khotanzad, A., & Hong, Y. H. (1990). Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 489–497.
- Khotanzad, A., & Hong, Y. H. (1991). Rotation invariant image recognition using features selected via a systematic

- method. *Pattern Recognition*, 23(10), 1089–1101.
- Khotanzad, A., & Lu, J. H. (1990). Classification of invariant image representation using a neural network. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(6), 1028–1038.
- Kim, W., & Park, R. H. (1990). Contour coding based on the decomposition of line segments. *Pattern Recognition Letters*, 11, 191–195.
- Lee, S. Y., Shan, M. K., & Yang, W. P. (1989). Similarity retrieval of iconic image database. *Pattern Recognition*, 22(6), 675–682.
- Lie, W. N., & Chen, Y. C. (1986). *Shape representation and matching using polar signature*. Paper presented at the International Computer Symposium, Tainan, Taiwan.
- Lin, C. S., & Hwang, C. L. (1987). New forms of shape invariants from elliptic Fourier descriptors. *Pattern Recognition*, 20(5), 535–545.
- Lin, W. G., & Wang, S. S. (1994). A note on the calculation of moments. *Pattern Recognition Letters*, 15, 1065–1070.
- Maitra, S. (1979). Moment invariants. *Proceedings of the IEEE*, 67, 697–699.
- Marshall, S. (1989). Review of shape coding techniques. *Image and Vision Computing*, 7(4), 281–294.
- Mehltre, B. M., & Narasimhalu, A. D. (1992). *An integrated shape and text retrieval system*. Technical Report, Institute of Systems Science, National University of Singapore, TR93-116-0.
- Mehltre, B. M., Kankanhalli, M. S., Man, G. C., & Narasimhalu, A. D. (1995). Color matching for image retrieval. *Pattern Recognition Letters*, 16, 325–331.
- Mustafa, Y. S. A., & Psaltis, D. (1984). Recognitive aspects of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 698–706.
- Nadler, M., & Smith, E. P. (1992). *Pattern recognition engineering*. New York: John Wiley & Sons.
- Pavlidis, T. (1980). Algorithms for shape analysis of contour and waveforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4).
- Pernus, F., Leonardis, A., & Kovacic, S. (1992). *Non-information preserving shape features at multiple resolution*. Paper presented at the 11th IAPR International Conference on Pattern Recognition.
- Persoon, E., & Fu, K. S. (1977). Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man and Cybernetics*, 7, 170–179.
- Rabitti, F., & Savino, P. (1991). *Automatic image indexation and retrieval*. Paper presented at the RIAO Conference on Intelligent Text and Image Handling.
- Rauber, T. W. (1994). Two-dimensional shape description. *Technical Report GR UNINOVA-RT-10-94, Universidade Nova de Lisboa, Lisboa, Portugal*.
- Rauber, T. W., & Steiger-Garçon, A. S. (1992). *Shape description by UNL Fourier features—an application to handwritten character recognition*. Paper presented at the 11th IAPR International Conference on Pattern Recognition, The Hague, The Netherlands.
- Raveendran, P., & Omatu, S. (1993). Performance of an optimal subset of Zernike features for pattern classification. *Information Sciences*, 1, 133–147.
- Reiss, T. H. (1991). The revised fundamental theorem of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 830–834.
- Robertson, S. E., & Hancock-Beaulieu, M. M. (1992). On the evaluation of IR systems. *Information Processing and Management*, 28(4), 439–440.
- Rosenfeld, A., & Kak, A. (1982). *Digital picture processing*, (Vol. 2, Chapter 11). New York: Academic Press.
- Salton, G. (1992). The state of retrieval system evaluation. *Information Processing & Management*, 28(4), 441–449.
- Scassellati, B., Alexopoulos, S., & Flickner, M. (1994). *Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgements*. Paper presented at the SPIE Conference on Storage and Retrieval for Image and Video Databases II.
- Smith, F. W., & Wright, M. H. (1971). Automatic ship photo interpretation by the method of moments. *IEEE Transactions on Computers*, 20, 1089–1094.
- Teague, M. (1980). Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70, 920–930.
- Tegolo, D. (1994). *Shape analysis for image retrieval*. Paper presented at the SPIE Conference on Storage and Retrieval for Image and Video Databases II.
- Teh, C. H., & Chin, R. T. (1988). On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4), 513–596.
- Wu, J. K., Mehltre, B. M., Gao, Y. J., Lam, P. C., & Narasimhalu, A. D. (1994). STAR—A multimedia database system for trademark registration. In Witold Litwin and Tore Risch (Eds.), *Applications of databases*. Lecture Notes in Computer Science (Vol. 819, pp. 109–122). Berlin: Springer-Verlag.
- Wu, J. K., Narasimhalu, A. D., Mehltre, B. M., Lam, C. P., & Gao, Y. J. (1995). CORE: a content based retrieval engine for multimedia information systems. *ACM Multimedia Systems*, 3(1), 1, 25–41.