

# Sketch-Based Image Matching Using Angular Partitioning

Abdollah Chalechale, *Student Member, IEEE*, Golshah Naghdy, and Alfred Mertins, *Senior Member, IEEE*

**Abstract**—This paper presents a novel method for image similarity measure, where a hand-drawn rough black and white sketch is compared with an existing data base of full color images (art works and photographs). The proposed system creates ambient intelligence in terms of the evaluation of nonprecise, easy to input sketched information. The system can then provide the user with options of either retrieving similar images in the database or ranking the quality of the sketch against a given standard, i.e., the original image model. Alternatively, the inherent pattern-matching capability of the system can be utilized to allow detection of distortion in any given real time-image sequences in vision-driven ambient intelligence applications. The proposed method can cope with images containing several complex objects in an inhomogeneous background. Two abstract images are obtained using strong edges of the model image and the morphologically thinned outline of the sketched image. The angular-spatial distribution of pixels in the abstract images is then employed to extract new compact and effective features using the Fourier transform. The extracted features are rotation and scale invariant and robust against translation. Experimental results from seven different approaches confirm the efficacy of the proposed method in both the retrieval performance and the time required for feature extraction and search.

**Index Terms**—Angular partitioning, Fourier transform, image matching, invariant properties, nonprecise interface, sketched images.

## I. INTRODUCTION

THE RAPID expansion and ease of acquisition of images in digital form have resulted in the availability of extensive image databases. Swift utilization and manipulation of these databases depends greatly on the intuitive user interfaces facilitating search, matching, and retrieval. Traditional textual methods have been shown to be inefficient and insufficient for searching in visual data. Consequently, image and video content-based indexing and retrieval methods have been the central themes of many new research projects in recent years [1], [2].

Existing content-based image retrieval (CBIR) systems include QBIC [3], PicSOM [4], MetaSEEk [5], VisualSEEk [6], and Blobworld [7]. The MPEG-7 standard defines descriptors

derived from three main image content features: color, texture, and shape [8], [9]. VisualSEEk and the algorithm proposed by Di Sciascio *et al.* [10] consider the spatial object layout as a significant content feature complementing color and texture attributes.

User interaction is one of the most important aspects of any multimedia system. A simple intuitive user interface makes the system more attractive and applicable. In sketch-based image retrieval (SBIR), where the query example is a rough and simple black and white hand-drawn draft, color and texture lose their original ability to serve as content keys. Visual shape descriptors are useful in SBIR when the model and the query image contain only one object in a plain background [11]. In multiple-object scenes, object layout is a powerful tool, however, object extraction and segmentation costs and rotation variance are the major drawbacks.

Edge pixel neighborhood information (EPNI) is utilizing a neighborhood structure of the edge pixels to make an extended feature vector [12]. The vector is used efficiently for measuring the similarity between sketched queries and arbitrary model images. The semantic power of the method is examined in [13]. Although the method is scale and translation invariant, it does not exhibit the rotation-invariance property. Matusiak *et al.* [14] and Horace *et al.* [11] have reported using rough and simple hand-drawn shape as input query for SBIR. The approach in [14] is based on curvature scale space, which is computationally expensive and has been shown to be less efficient than Fourier descriptors and Zernike moments techniques [15]. In [11], several dominant points are extracted for each contour using information derived from the convex hull and the contour curvature. Query-by-sketch is adequately investigated in [10] using the spatial relationships between shapes in an image. The approach introduces a way to represent shapes, their spatial arrangement, and color and texture attributes in the user sketch. Furthermore, in this approach, segmentation of the database image is required and the sketched queries include color and texture attributes.

Intuitive user interfaces, such as sketch-based ones, liberate the user from concerns about precision, orientation, scale, texture and color. In this paper the focus, therefore, is on the problem of finding image features, invariant to rotation and scale changes, which can be used efficiently in sketch-based retrieval, where images have several objects in an inhomogeneous background. Since object extraction and segmentation are not needed in this approach, therefore, the images (model and query) may consist of several complex objects. The input query has no color and texture attributes. We also eliminate any constraint regarding the shape of the objects and the existence of any background.

Manuscript received October 13, 2003; revised March 25, 2004, June 1, 2004, and June 18, 2004. The work of A. Chalechale was supported by the Ministry of Science, Research, and Technology, I. R. Iran. This paper was recommended by Guest Editor G. L. Foresti.

A. Chalechale and G. Naghdy are with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, NSW 2522, Australia (e-mail: ac82@uow.edu.au; golshah@uow.edu.au).

A. Mertins is with the Signal Processing Group, Institute of Physics, University of Oldenburg, Oldenburg 26111, Germany (e-mail: alfred.mertins@uni-oldenburg.de).

Digital Object Identifier 10.1109/TSMCA.2004.838464

In this paper, we present a novel approach of feature extraction for matching purposes based on angular partitioning of two abstract images. Abstract images are obtained from the model image and from the query image by two different procedures. The angular-spatial distribution of pixels in the abstract image is then employed as the key concept for feature extraction using the Fourier transform. The extracted features are scale and rotation invariant and robust against translation.

The proposed system finds applications in sketch-based image retrieval, nonprecise human-machine interfaces, and can be adapted for artistic drawing-skill training. It creates ambient intelligence in terms of the evaluation of nonprecise, easy to input sketched information. The system can provide the user with options of either retrieving similar images in the database or ranking the quality of the sketch against a given standard, e.g., the original model image. The major contribution of the paper is in segmentation-free SBIR, with scale and rotation invariance properties. The proposed algorithm and six other well-known approaches within the literature are implemented and examined using an art and photograph image database (ART-PHOTO BANK). The aim is to show the degree of rotation and scale invariance separately. Experimental results confirm the superiority of the proposed method using the MPEG-7 standard retrieval performance assessment measure, known as average normalized modified retrieval rank (ANMRR), and computational speed measures.

The outline of the paper is as follows. In Section II, we explain related work, which can be used or adapted for SBIR. The details of the proposed approach are discussed in Section III. Section IV exhibits evaluation criteria and detailed experimental results. It also provides a comparison of the proposed method with six other approaches within the literature based on the retrieval performance, feature extraction, and searching times (SETs). Section V concludes the paper and poses some new directions.

## II. RELATED WORK

While the general problem of CBIR has received a lot of attention, not a great deal of works address SBIR, which plays an important role in intelligent systems developed for nonprecise human machine interface. In the following, we briefly describe some approaches which can be used directly or can be adapted for SBIR.

The work of Hirata and Kato, query by visual example (QVE) [16], is one of the earliest approaches that addresses SBIR. The IBM Corporation adapts a modified version of the approach in its QBIC system [3]. It defines a pictorial index for each image, including query and database images. The QVE method performs retrieval by computing the correlation between the corresponding indexes. In this approach, the query and the database images are resized to  $64 \times 64$  pixels and then a proposed gradient operator extracts their edges. The resulting edge maps are called pictorial indexes and are used for image-to-image matching. After dividing each pictorial index image into 64 blocks of equal size, the correlation between

corresponding blocks in the query index  $q$  and the database index  $p$  is calculated by the following bit-wise summation with shifting blocks of  $q$  by  $\delta$  and  $\varepsilon$  over blocks of  $p$ :

$$C_{\delta\varepsilon} = \sum_s \sum_r (\alpha p_{rs} \cdot q_{r+\delta s+\varepsilon} + \beta \overline{p_{rs}} \cdot \overline{q_{r+\delta s+\varepsilon}} + \gamma p_{rs} \oplus q_{r+\delta s+\varepsilon}). \quad (1)$$

The coefficients  $\alpha$ ,  $\beta$ , and  $\gamma$  are control parameters used to estimate matching and mismatching patterns and take up values 10, 1, and  $-3$ , respectively. The maximum value of  $C_{\delta\varepsilon}$  for all  $\delta$  and  $\varepsilon$  for each block is called the local correlation factor  $C_L$

$$C_L = \max(C_{\delta\varepsilon}), \quad \text{for } -4 \leq \delta \leq 4 \quad \text{and} \quad -4 \leq \varepsilon \leq 4. \quad (2)$$

Finally, a global correlation factor  $C_G$  that is the sum of all 64  $C_L$ 's is calculated and used as the similarity measure.

Although the method has a good ability to find similar images in small data sets, it does not allow efficient indexing [16]. Furthermore, because of the expensive computational cost it is time consuming to use the method in a large image database. While the method can tolerate minor local rotations, it is not rotation invariant and does not allow for large global rotations.

The histogram of edge directions (HED), representing image information, is one of the well-known methods in the image retrieval literature [17]–[19]. Abdel-Mottaleb [20] utilizes this method by applying the Canny edge operator [21] to find strong edges in an image and then quantizes them into four directions (horizontal, vertical, and the two diagonals) to build histograms of edge directions for different image regions. The histograms are then used as hash values in a hash table indexing scheme. Jain and Vailaya [22] use edge directions as an image attribute for shape description. They show that in the absence of color information or in images with homogenous colors this histogram is a significant tool in searching for similar images. They also exploit the histogram in conjunction with invariant moments in a case study of trademark registration process [23]. The edge information contained in the database images is extracted off-line using the Canny edge operator. The corresponding edge directions are, subsequently, quantized into 72 bins of  $5^\circ$  each. To reduce the effect of rotation, they smooth the histogram as follows:

$$I_s[i] = \frac{\sum_{j=i-k}^{i+k} I[j]}{2k+1} \quad (3)$$

where  $I_s$  is the smoothed histogram,  $I$  is the original normalized histogram, and the parameter  $k$  determines the degree of smoothing.

Moment invariants are widely used in pattern recognition and image analysis [23], [24], [25]. Geometric moments are defined as

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (4)$$

where  $p, q = 0, 1, 2, \dots, \infty$ , and  $f(x, y)$  is the density distribution function of the image. Seven invariant moment functions ( $\phi_1 - \phi_7$ ) are defined based on normalized central geometric moments [26]. The first six functions ( $\phi_1 - \phi_6$ ) are

invariant under rotation, and the last one  $\phi_7$  is both skew and rotation invariant.

Zernike moments, on the other hand, are less sensitive to noise than geometric moments and are more powerful in discriminating objects with  $n$ -fold symmetries [27]. They are exploited for building a region-based shape descriptor in the MPEG-7 standard. Zernike orthogonal polynomials are employed to derive Zernike moment invariants (ZMI) of an image  $f(x, y)$  as follows:

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^\infty [V_{nl}(r, \theta)]^* f(r \cos \theta, r \sin \theta) r dr d\theta \quad (5)$$

where  $n = 0, 1, 2, \dots, \infty$ , and  $l$  takes on positive and negative integer values subject to the condition  $n - |l|$  even and  $|l| \leq n$ . The Zernike polynomial is defined as

$$V_{nl}(x, y) = R_{nl}(r) e^{il\theta} \quad (6)$$

and the radial polynomial is

$$\begin{aligned} R_{nl}(r) &= \sum_{s=0}^{(n-|l|)/2} (-1)^s \\ &\cdot \frac{(n-s)!}{s!((n+|l|)/2-s)!((n-|l|)/2-s)!} r^{n-2s} \\ &= \sum_{\substack{k=|l| \\ n-k \text{ even}}}^n B_{n|l|k} r^k. \end{aligned} \quad (7)$$

If only Zernike moments of order less than or equal to  $N$  are given, then the image function  $f(x, y)$  can be approximated by

$$f(x, y) \approx \sum_{n=0}^N \sum_{\substack{l \\ n-|l| \text{ even}, |l| \leq n}} A_{nl} V_{nl}(x, y). \quad (8)$$

The magnitudes of  $\{A_{nl}\}$  are used for image matching and the similarity between images is measured using the  $\ell_1$  (Manhattan) distance [24], [27]. In the case of hand-drawn sketched images,  $f(x, y)$  is a binary image representing the outline of the query. One obvious advantage of the use of binary functions is the low computational complexity. As  $f(x, y)$  is a black and white sketch, computation is not required for every pixel. In fact, computation is dependent on the number of pixels in the sketched query. This approach has been employed for SBIR in [27] and shows better performance than the traditional two-dimensional (2-D) Fourier transform method.

The MPEG-7 standard defines the edge histogram descriptor (EHD) for texture characterization [28]. The distribution of edges is not only a good texture signature, it is also useful for image-to-image matching in the absence of any homogeneous texture. This descriptor has been used for SBIR in the literature [29]. A given image is first divided into 16 subimages ( $4 \times 4$ ), and local edge histograms are computed for each subimage. To compute the edge histogram, each of the 16 subimages is further subdivided into image blocks. The size of each image block is proportional to the size of the original image and is assumed to be a multiple of two. The number of image

blocks, independent of the original image size, is constant ( $desired\_num\_of\_blocks$ ) and the block size is figured as follows:

$$x = \sqrt{\frac{image\_width * image\_height}{desired\_num\_of\_blocks}} \quad (9)$$

$$block\_size = \left\lfloor \frac{x}{2} \right\rfloor * 2 \quad (10)$$

where  $image\_width$  and  $image\_height$  represent the horizontal and vertical size of the image, respectively. Each image block is then partitioned into four ( $2 \times 2$ ) blocks of pixels and the pixel intensities, for these four divisions are computed by averaging the luminance values of the existing pixels. In the case of black and white sketch images, the luminance takes only the value of one or zero. Edges are grouped into only five classes: vertical, horizontal,  $45^\circ$  diagonal,  $135^\circ$  diagonal, and isotropic (nondirectional), based on directional edge strengths. These directions are determined for each image block using five corresponding  $2 \times 2$  filter masks corresponding to  $2 \times 2$  subdivisions of the image blocks. If the maximum directional strength is greater than a threshold value ( $Th_{edge}$ ) then the underlying block is designated to belong to the corresponding edge class. The default value of ( $Th_{edge}$ ) for gray scale images is 11 and for binary sketches it is set to 0.

The histogram for each subimage represents the frequency of occurrence of the five classes of edges in the corresponding subimage. As there are 16 subimages and each has a 5-bin histogram, a total of  $16 \times 5 = 80$  bins in the histogram is achieved. For normalization, the number of edge occurrences for each bin is divided by the total number of image blocks in the subimage. To minimize the overall number of bits, the normalized bins are nonlinearly quantized and fixed-length coded with three bits per bin, resulting in a descriptor of size 240 bits (see [28] for the details of the quantization process).

Won *et al.* [30] proposed efficient use of the EHD by extending the histogram to 150 bins. The extended histogram is obtained by grouping the image blocks into 13 clusters (four vertical, four horizontal, and five square clusters). Each cluster contains four subimages. In addition to this semiglobal histogram with  $13 \times 5 = 65$  bins, another 5-bin global histogram is computed by combining all 16 local bins. This results in a 150 ( $80+65+5$ ) bin histogram that is used for measuring the similarity between images. The global and the semiglobal histograms could be reconstructed directly from the local histogram at the matching time. Although the approach achieves good retrieval performances in some applications [28], [30], experimental results (Section IV) show that it does not exhibit the rotation invariance property since it only considers five predefined edge directions.

The angular radial transform (ART) descriptor is an efficient tool to retrieve object information [9]. An ART-based descriptor is adopted by MPEG-7 [28] as shape descriptor. The descriptor can describe complex objects (consisting of multiple disconnected regions, such as trademarks) as well as simple objects. As the descriptor is based on the regional property, it shows robustness to segmentation noise, e.g., salt and pepper noise.

By definition, the ART is a unitary transform defined on a unit disk that consists of the complete orthonormal sinusoidal

basis functions in polar coordinates. From each image, a set of ART coefficients  $\psi_{mn}$  of order  $m$  and  $n$  is extracted using the following formula:

$$\psi_{mn} = \int_0^{2\pi} \int_0^1 V_{mn}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta \quad (11)$$

where  $f(\rho, \theta)$  is an image intensity function in polar coordinates, and  $V_{mn}(\rho, \theta)$  is the ART basis function of order  $m$  and  $n$  that is separable along the angular and radial directions, i.e.,

$$V_{mn}(\rho, \theta) = A_n(\theta) R_m(\rho). \quad (12)$$

In order to achieve rotation invariance, an exponential function is used for the angular basis function

$$A_n(\theta) = \frac{1}{2\pi} e^{jn\theta} \quad (13)$$

and the radial basis function defined by a cosine function

$$R_m(\rho) = \begin{cases} 1, & m = 0 \\ 2 \cos(\pi m \rho), & m \neq 0 \end{cases}. \quad (14)$$

It can be shown that the magnitudes of the ART are rotation invariant [31]. The discrete ART coefficients of a gray scale image can be found easily using a look-up table. Size normalization, ART transformation, and area normalization are applied consecutively. In the first step, the size of the image is normalized by linear interpolation to a predefined width and height ( $101 \times 101$ ). An edge detector algorithm such as the Canny edge operator is subsequently used to obtain the size invariant edge map  $I(x, y)$ . The ART transformation is then applied to the edge map. Finally, dividing each magnitude of the ART coefficients by the magnitude of the first coefficient (i.e.,  $\|\psi_{00}\|$ ) yields normalized ART coefficients that will be used for the similarity measure.

The 2-D Fourier transform in polar coordinates is employed for shape description in [32]. Its supremacy over the one-dimensional (1-D) Fourier descriptors, curvature scale space descriptors and the Zernike moments has been shown in [15]. In this approach, the polar shape image is treated as a normal rectangular image and the 2-D Fourier transform is applied on this rectangular image. This polar Fourier transform has a form that is similar to the normal 2-D discrete Fourier transform in Cartesian coordinates. Consequently, for a given shape image  $f(x, y)$ , the PFT descriptor is obtained as follows:

$$pft(\rho, \phi) = \sum_r \sum_i f(r, \theta_i) e^{[j2\pi(\frac{\rho}{R}\rho + \frac{2\pi i}{T}\phi)]} \quad (15)$$

where  $r = [(x - x_c)^2 + (y - y_c)^2]^{0.5}$ , ( $0 \leq r < R$ ), and  $\theta_i = i(2\pi/T)$ , ( $0 \leq i < T$ );  $0 \leq \rho < R$ ,  $0 \leq \phi < T$ .  $(x_c, y_c)$  is the center of mass of the shape, and  $R$  and  $T$  are the radial and angular resolutions.  $\rho$  and  $\phi$  are the  $\rho$ th radial frequency and the  $\phi$ th angular frequency, respectively, selected for image description. The above extracted Fourier coefficients are only translation invariant [32]. The following normalization process makes them scale and rotation invariant [33]:

$$\text{PFD} = \left\{ \frac{|pft(0, 0)|}{area}, \frac{|pft(0, 1)|}{|pft(0, 0)|}, \dots, \frac{|pft(0, n)|}{|pft(0, 0)|}, \dots, \frac{|pft(m, 0)|}{|pft(0, 0)|}, \dots, \frac{|pft(m, n)|}{|pft(0, 0)|} \right\} \quad (16)$$

where  $area$  is the area of the bounding circle in which the polar image resides.  $n$  is the maximum number of angular frequencies and  $m$  is the maximum number of radial frequencies selected.

### III. ANGULAR PARTITIONING OF ABSTRACT IMAGE (APAI)

The main objective of the method proposed in this paper is to transform the image data into a new structure that supports measuring the similarity between a full color image and a black and white hand-drawn simple sketch.

The edge map of an image carries the solid structure of the image independent of the color attribute. Edges are also proven to be a fundamental primitive of an image for the preservation of both semantics and perceived attributes [34]. Furthermore, in SBIR, edges form the most useful feature for matching purposes [11], [12], [14]. According to the assumption that sketched queries are more similar to the edge maps, which contain only the perceptive and vigorous edges, we obtain two abstract images through the strong edges of the model image and the thinned version of the query image. The proposed features are then extracted from the abstract images.

#### A. Image Abstraction

The full-color model image is initially converted to a gray intensity image  $I$  by eliminating the hue and saturation while retaining the luminance. The edges are then extracted using the Canny operator with  $\sigma = 1$  and Gaussian mask of size = 9 using the following procedure for depicting the most perceived edges. The values of high and low thresholds for the magnitude of the potential edge points are automatically computed in such a way that only the strong edges are retained. This improves the general resemblance of the resulting edge map and the hand-drawn query. In order to depict strong edges, let  $G$  be the Gaussian 1-D filter and let  $g$  be the derivative of the Gaussian used in the Canny edge operator. Then,

$$H(k) = \sum_i G(i)g(k+1-i) \quad (17)$$

is the 1-D convolution of the Gaussian and its derivative.

$$X(u, v) = \left[ \sum_{j=1}^V I'(u, j)H(v-j) \right]' \quad (18)$$

and

$$Y(u, v) = \sum_{i=1}^U I(i, v)H(u-i) \quad (19)$$

where  $u = 1, 2, 3, \dots, U$  and  $v = 1, 2, 3, \dots, V$  are the vertical and horizontal edge maps, respectively.  $U$  is the number of rows and  $V$  is the number of columns in the image  $I$ . The  $'$  notation indicates matrix transpose. The magnitude of the edge points is then obtained as

$$\Gamma(u, v) = \sqrt{X(u, v)^2 + Y(u, v)^2}. \quad (20)$$

For efficient selection of the high and low thresholds, we make a 64-bin cumulative histogram of the  $\Gamma(u, v)$  values and find the minimum index  $\iota$  in this cumulative histogram that is greater than  $\alpha \cdot U \cdot V$ , where  $\alpha$  denotes the percentage of nonedge points in the image ( $\alpha = 0.7$  is an adequate choice for many



Fig. 1. Effect of  $\beta$  parameter on edge maps. (a) Model image, (b)  $\beta = 1$ , (c)  $\beta = 2$ , (d)  $\beta = 3$ .

images). To retain strong edges of the image,  $\beta \cdot \iota$  is selected as the high-threshold value and  $0.4\beta \cdot \iota$  is used for the low-threshold value in the Canny edge operator.  $\beta$  is a parameter that controls the degree of the strength of the edge points. Higher  $\beta$ 's lead to a lower number of edge points, but more perceptible ones (see Fig. 1). Consequently, the gray image  $I$  is converted to an edge image  $P$ , using the Canny edge operator exploiting the above automatic extracted thresholds.

For the query images, black and white morphological thinning [35] is applied to extract a thinned version of the sketched image. This image, denoted as  $Q$ , shows an outline of the query and contains the main structure of the user request. It contains the spatial distribution of pixels similar to the strong edge map of the model image  $P$ .

To achieve scale invariance and robustness against translations in the space of query and model images, the following normalization procedure is applied. First, the bounding boxes of the images  $P$  and  $Q$  are obtained. The area in the bounding box of the comparable images are then normalized to  $J \times J$  pixels using a nearest neighbor interpolation. The proposed normalization of the images  $P$  and  $Q$  then ensures scale invariance since different size images are converted to predefined-size images. Working in the bounding box of images lets the input sketch to be located anywhere in the scanned paper. The resulting image is called *abstract image*  $\Omega$ . It is the input to the following partitioning procedure.

### B. Angular Partitioning

We define angular partitions (slices) in the surrounding circle of the abstract image  $\Omega$  (see Fig. 2). The angle between adjacent slices is  $\varphi = 2\pi/K$ , where  $K$  is the number of angular partitions in the abstract image.  $K$  can be adjusted to achieve hierarchical coarse to fine representations. Any rotation of a given image, with respect to its center, moves a pixel at slice  $S_i$  to a new position at slice  $S_j$ , where  $j = (i + \lambda) \bmod K$ , for  $i, \lambda = 0, 1, 2, \dots, K-1$ .

The number of edge points in each slice of  $\Omega$  is chosen to represent the slice feature. The scale invariant image feature is then  $\{f(i)\}$ , where

$$f(i) = \sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^R \Omega(\rho, \theta) \quad (21)$$

for  $i = 0, 1, 2, \dots, K-1$ .  $R$  is the radius of the surrounding circle of the abstract image. The feature extracted above will be circularly shifted when the image  $\Omega$  is rotated  $\tau = l2\pi/K$  radians

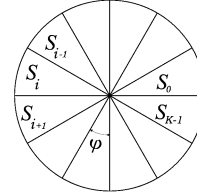


Fig. 2. Angular partitioning splits the abstract image into  $K$  successive slices.

( $l = 0, 1, 2, \dots$ ). To show this, let  $\Omega_\tau$  denote the abstract image  $\Omega$  after rotation by  $\tau$  radians in counterclockwise direction

$$\Omega_\tau(\rho, \theta) = \Omega(\rho, \theta - \tau). \quad (22)$$

Then,

$$f_\tau(i) = \sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^R \Omega_\tau(\rho, \theta) \quad (23)$$

are the image feature elements of the  $\Omega_\tau$  for the same  $i$ 's. We can express  $f_\tau$  as

$$\begin{aligned} f_\tau(i) &= \sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^R \Omega(\rho, \theta - \tau) \\ &= \sum_{\theta=\frac{(i-l)2\pi}{K}}^{\frac{(i-l+1)2\pi}{K}} \sum_{\rho=0}^R \Omega(\rho, \theta) \\ &= f(i-l) \end{aligned} \quad (24)$$

where  $i-l$  is a modulo  $M$  subtraction.  $f_\tau(i) = f(i-l)$  indicates that there is a circular shift in the image feature  $\{f_\tau(i)\}$  corresponding to the image feature  $\{f(i)\}$ , representing  $\Omega_\tau$  and  $\Omega$ , respectively.

Using the 1-D discrete Fourier transform of  $f(i)$  and  $f_\tau(i)$  we obtain

$$\begin{aligned} F(u) &= \frac{1}{K} \sum_{i=0}^{K-1} f(i) e^{-j2\pi ui/K} \\ F_\tau(u) &= \frac{1}{K} \sum_{i=0}^{K-1} f_\tau(i) e^{-j2\pi ui/K} \\ &= \frac{1}{K} \sum_{i=0}^{K-1} f(i-l) e^{-j2\pi ui/K} \\ &= \frac{1}{K} \sum_{i=-l}^{K-1-l} f(i) e^{-j2\pi u(i+l)/K} \\ &= e^{-j2\pi ul/K} F(u). \end{aligned} \quad (25)$$

Based on the property  $|F(u)| = |F_\tau(u)|$ , the scale and rotation invariant image features are chosen as  $\Upsilon = \{|F(u)|\}$  for  $u = 0, 1, 2 \dots K - 1$ . The extracted features are robust against translation because of the aforementioned normalization process. Choosing a medium-size slice (e.g.,  $\varphi = 30^\circ$ ) makes the extracted features more vigorous against local variations which are common in hand-drawn sketches. This is based on the fact that the number of pixels in such slices varies slowly with local translations. The features are rotation invariant due to the Fourier transform applied.

Fig. 3 shows an image example, its  $90^\circ$  rotated version, corresponding abstract images before size normalization, abstract images superimposed with the APAI slices, and extracted  $\{f(i)\}$  and  $\Upsilon = \{|F(u)|\}$  features. Experimental results (Section IV) confirm the robustness and efficiency of the proposed method.

#### IV. EXPERIMENTAL RESULTS AND EFFICACY EVALUATION

In this section, first, the similarity measurement and criteria employed for retrieval performance assessment are discussed. Then, the effect of different parameters of the APAI method is investigated and finally comparative results, showing the degree of rotation and scale invariant properties, are presented.

##### A. Similarity Measurement and Retrieval Performance Criteria

The similarity between images is measured by the  $\ell_1$  (Manhattan) distance between the two corresponding feature vectors. Suppose  $\Upsilon_P$  and  $\Upsilon_Q$  represent two different images  $P$  and  $Q$  respectively. The similarity between image  $P$  and  $Q$  is the inverse of their Manhattan distance calculated as

$$d(P, Q) = \sum_u |\Upsilon_P(u) - \Upsilon_Q(u)|. \quad (26)$$

*Recall* and *precision* are well-known retrieval performance measures. They are basically hit-and-miss counters. In other words, the retrieval performance is based on the number of retrieved images which have similarity measures that are greater than a given threshold. For more specific comparisons, however, we also need the rank information among the retrieved images. The ANMRR, which was developed during the MPEG-7 standardization activity, is a measure that exploits not only the Recall and Precision information but also the rank information among the retrieved images. It is defined in the MPEG-7 standard [36] as follows:

$$\text{AVR}(q) = \sum_{k=1}^{\text{NG}(q)} \frac{\text{Rank}(k)}{\text{NG}(q)} \quad (27)$$

$$\text{MRR}(q) = \text{AVR}(q) - 0.5 - \frac{\text{NG}(q)}{2} \quad (28)$$

$$\text{NMRR}(q) = \frac{\text{MRR}(q)}{L + 0.5 - 0.5 * \text{NG}(q)} \quad (29)$$

$$\text{ANMRR} = \frac{1}{Q_{\text{tot}}} \sum_{q=1}^{Q_{\text{tot}}} \text{NMRR}(q) \quad (30)$$

where  $\text{NG}(q)$  is the number of ground truth images for a query  $q$ .  $L = \min(4 * \text{NG}(q), 2 * \text{GTM})$ , where  $\text{GTM} = \max\{\text{NG}(q)\}$  for all  $q$ 's of a data set.  $\text{Rank}(k)$  is the rank of the found ground

truth images, counting the rank of the first retrieved image as one. A Rank of  $L + 1$  is assigned to each of the ground truth images which are not in the first  $L$  retrievals.  $Q_{\text{tot}}$  is the total number of queries in the test. For example, suppose a given query  $q_i$  has ten similar images in an image database ( $\text{NG} = 10$ ). If an algorithm finds six of them in the top 20 retrievals ( $L = 20$ ) in the ranks of 1, 5, 8, 13, 14, and 18, then the  $\text{AVR}(q_i) = 14.3$ ,  $\text{MRR}(q_i) = 8.8$ , and  $\text{NMRR}(q_i) = 0.5677$ . Note that the  $\text{NMRR}$  and its average ( $\text{ANMRR}$ ) will always be in the range of  $[0, 1]$ . Based on the definition of  $\text{ANMRR}$ , the smaller the  $\text{ANMRR}$ , the better the retrieval performance.

The receiver operating characteristic (ROC) curve [37], an interesting analysis tool in two-class problems, is employed to evaluate the ability of the system when no similar images to a given query exist. To this end, queries with no similar images in the database, in addition to queries with some similar images in the database, are given to the system and the response of the system which could be either 1) "there exist some similar images" or 2) "there exist no similar image" is logged. For different threshold values, the true positive ratio [(TPR) *sensitivity*] and the false positive ratio [(FPR) 1-*specificity*] are computed [38] and the corresponding ROC curve is depicted.

One of the difficulties involved in achieving an ROC curve for SBIR is the diversity in the range of the similarity measures. To select an effective threshold for all queries, their distances to the database images should be in a tantamount interval. In order to overcome the problem, distance values are normalized to be within the same range of  $[0, 1]$ . More precisely, let  $d(P, Q)$  denote the Manhattan distance between  $Q$  and  $P$  images. The normalization is done as follows:

$$d_n(P, Q) = \frac{d(P, Q) - \text{mindist}}{\text{maxdist} - \text{mindist}} \quad (31)$$

where  $d_n(P, Q)$  is the normalized distance. *mindist* and *maxdist* are the minimum and the maximum distance values of the query image  $Q$  to the database images according to the corresponding feature vectors. Accordingly, the threshold for computing *sensitivity* and *specificity* varies in the interval of zero to one for all queries.

##### B. Effect of Parameter Variations

In order to evaluate retrieval effectiveness of the proposed method and to test the effects of parameter variations on the retrieval performance, a database of model and query images was created and several experiments were conducted. The database is a collection of different model and query images called ART-PHOTO BANK. Currently, it contains 4000 full color heterogeneous images of various sizes in the model part (500 in groups of eight) and 400 sketches in its query part (100 in groups of four). Images in the model part are a true-balanced combination of 250 art works, gained from the World Art Kiosk, California State University, and 250 real natural photographs from set S3 of the MPEG-7 database. Each group contains eight similar images created by rotation in steps of  $45^\circ$ . This results in a variety of scaled and rotated samples. Images in the query part are hand-drawn black and white sketches similar to 100 arbitrary candidates from the model part and their rotated versions ( $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ). This is to simulate different vertical and

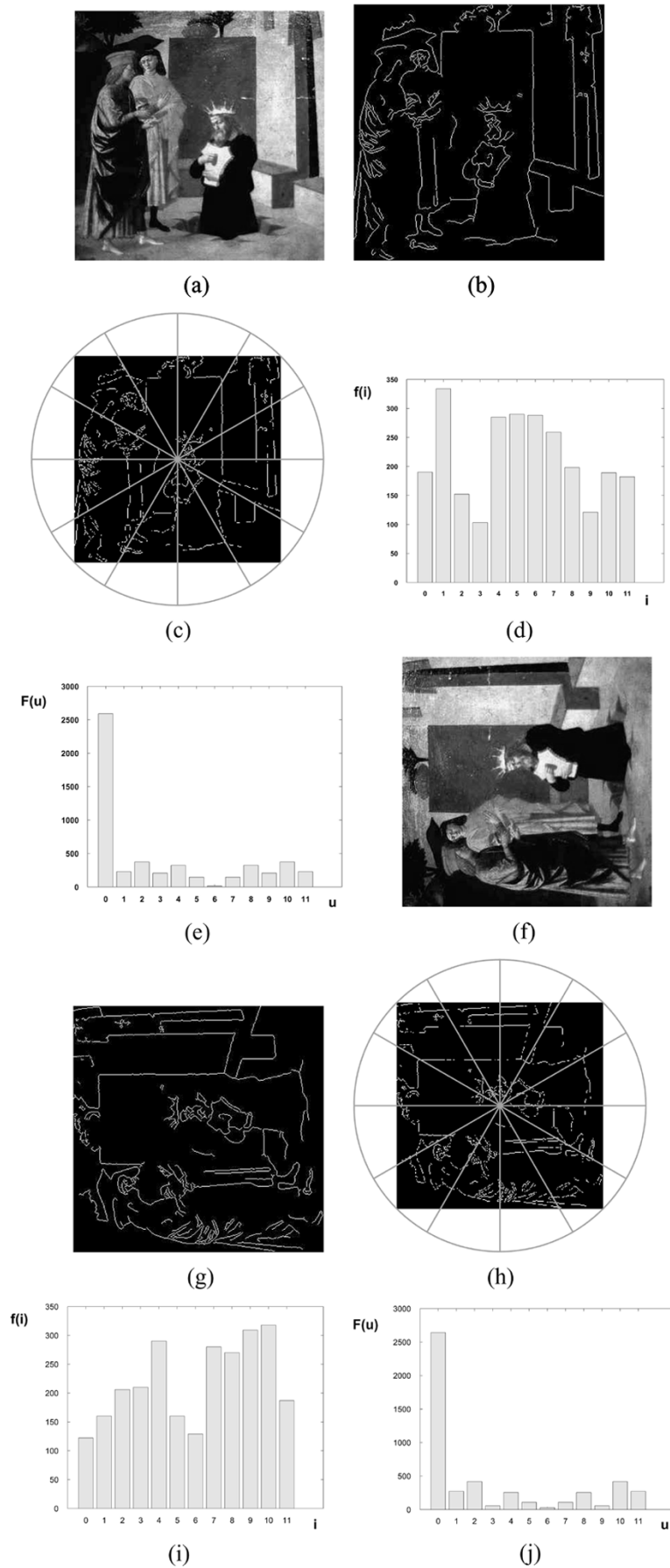
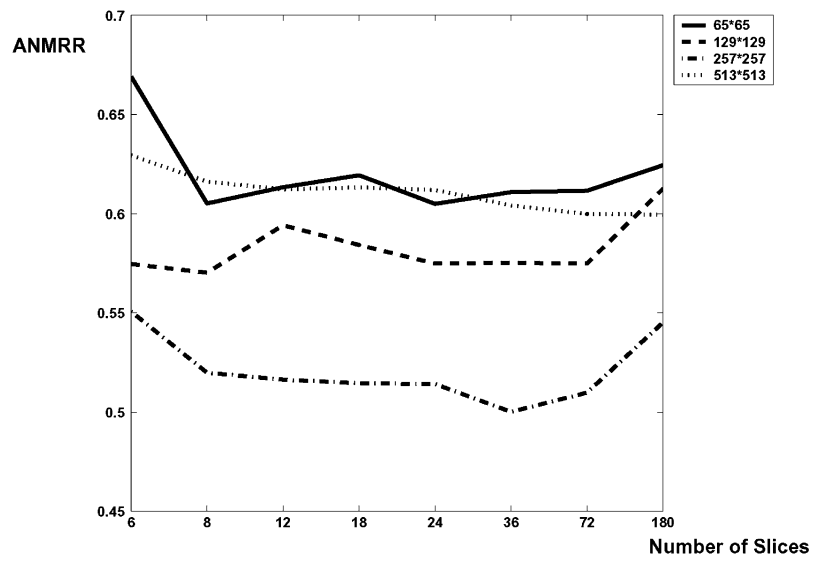
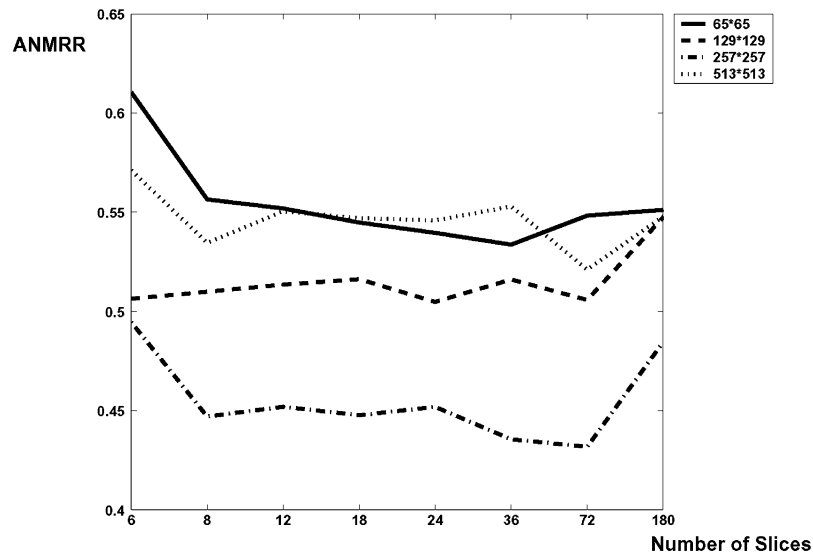


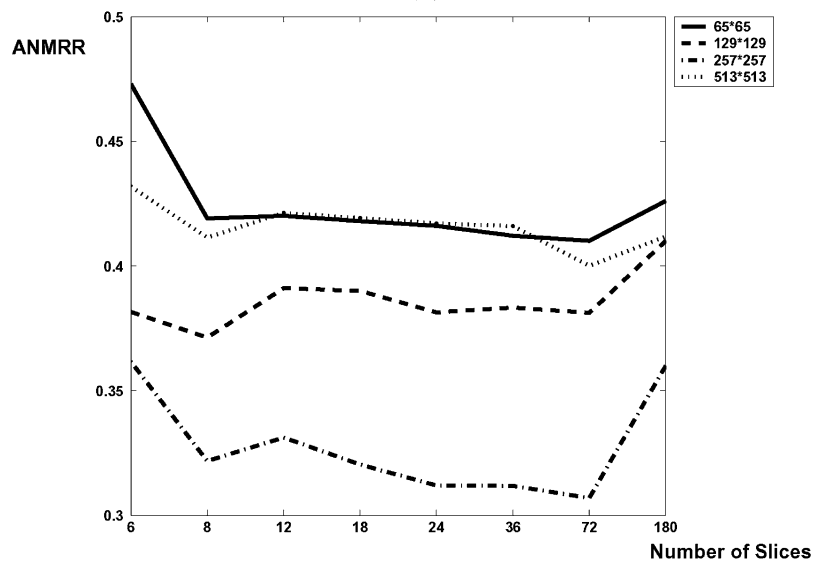
Fig. 3. (a) Image example and (f) the  $90^\circ$  rotated version. (b) and (g) Corresponding abstract images before size normalization. (c) and (h) Abstract images superimposed with angular partitions. (d) and (i) Number of pixels in different slices. (e) and (j) Invariant features extracted using the Fourier transform.



(a)



(b)



(c)

Fig. 4. ANMRR measure of the APAI method with (a)  $\beta = 1$ , (b)  $\beta = 2$ , and (c)  $\beta = 3$  for a varying number of slices and four different normalized sizes.



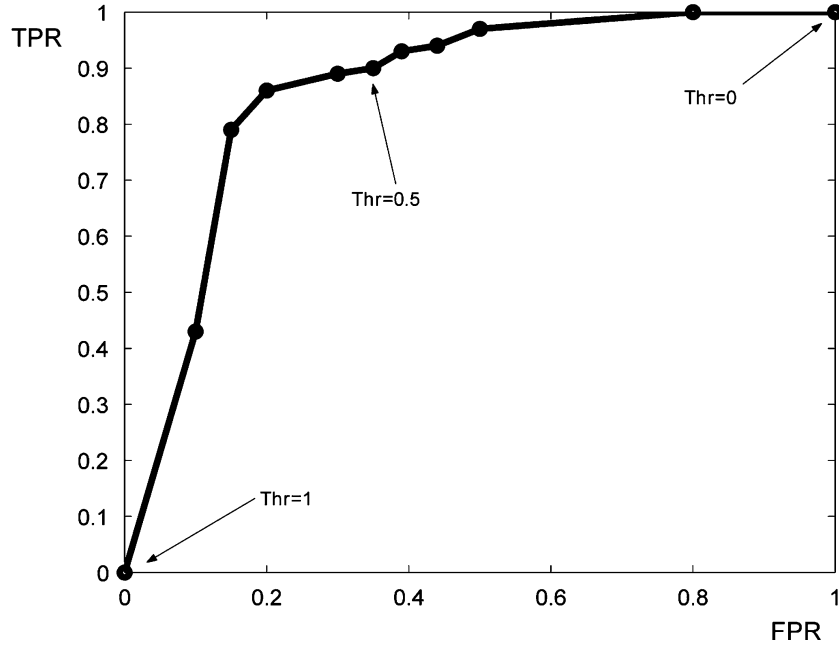


Fig. 5. ROC curve of the APAI method showing the TPR and the FPR for different thresholds. The threshold (Thr) is applied on the normalized distance  $d_n(P, Q)$ .

horizontal directions when posing a sketched query to the retrieval system. Sketches were drawn by different users and then scanned with 200 dpi resolution. See Figs. 7–9 for some examples of the sketched and the model images. Note that here the ANMRR picks up the 16 ( $2 \times \text{GTM}$ ) top similar images. Therefore, as each input query has eight (GTM) similar images in the database, in the best case there are eight nonsimilar images in the retrieval list.

The performance of the proposed method for various internal parameters is evaluated using ANMRR measure. Different levels of abstraction ( $\beta$  parameter) and different numbers of angular partitions ( $\varphi$  parameter) for several size normalization parameter  $J$  were tested. This is to determine which set is the most suitable for SBIR. In our experiments we have chosen  $NG(q) = 8$  for all  $qs$ ,  $L = 16$ , and  $Q_{\text{tot}} = 400$ .

Fig. 4(a)–(c) depict the resulting ANMRR for  $\beta = 1, 2$ , and  $3m$  respectively. In each part, four different size normalization parameters  $J$  with eight different numbers of angular partitions (corresponding to  $\varphi = 2^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 30^\circ, 45^\circ$ , and  $60^\circ$ ) were examined. Fig. 4(c), which relates to  $\beta = 3$ , exhibits the best results [generally, Fig. 4(c) exhibits smaller ANMRR values than (a) and (b)]. Higher  $\beta$ s were also examined (not reported here), but the results were worse than those in Fig. 4(b). The optimum performance for  $\beta = 3$  is due to the following reasons: 1) smaller  $\beta$ s generate significant noise in the edge images resulting in considerable variation among the abstract images (see Fig. 1) and 2) higher  $\beta$ s generate edge images, whereas essential information is lost, rendering the comparison with query images ineffective.

Normalizing the image size to  $257 \times 257$  yields the best results. This is based on the fact that images in the database having an average size closer to  $257 \times 257$  than the other sizes ( $65 \times 65$ ,  $129 \times 129$ , and  $513 \times 513$ ). Consequently, normalizing with this size keeps more information in the abstract images and reduces the adverse effect of normalization.

With smaller  $\varphi$ s (higher numbers of slices), the proposed method exhibits better performance with the exception of  $\varphi = 2^\circ$  [see Fig. 4(c)]. The reason is that the system can capture more details with a medium to small number of slices, but with very small slices, the method loses the robustness against small translations which always exist between abstract images and the tinned sketched queries and consequently the overall performance degrades. Therefore,  $\beta = 3$ ,  $J = 257$ , and  $\varphi = 5^\circ$  are chosen as the system's internal characteristics.

Fig. 5 shows the ROC curve of the system with the aforementioned parameters. The false positive ratio (FPR) i.e.,  $1 - \text{specificity}$ , is obtained, posing 50 queries which have some similar images in the database, using

$$\text{FPR} = \frac{C}{C + D} \quad (32)$$

where  $C$  is the number of cases where the system says “there exist no similar image,” and  $D$  is the number of cases where the system says “there exist some similar images”. Similarly, the TPR, i.e., *sensitivity*, is obtained posing another 50 queries, which have no similar images in the database, using

$$\text{TPR} = \frac{A}{A + B} \quad (33)$$

where  $A$  is the number of cases where the system says “there exist no similar image,” and  $B$  is the number of cases where the system says “there exist some similar images”. Threshold values are set to 1 (lower-left corner) downward to zero (upper-right corner) with the step of 0.1 on the normalized distance  $d_n(P, Q)$ . As it can be seen, the *sensitivity* of the proposed method is higher than its *specificity*. This comes from the fact that rejecting queries with no similar images is easier than finding similar images to a given sketched query.

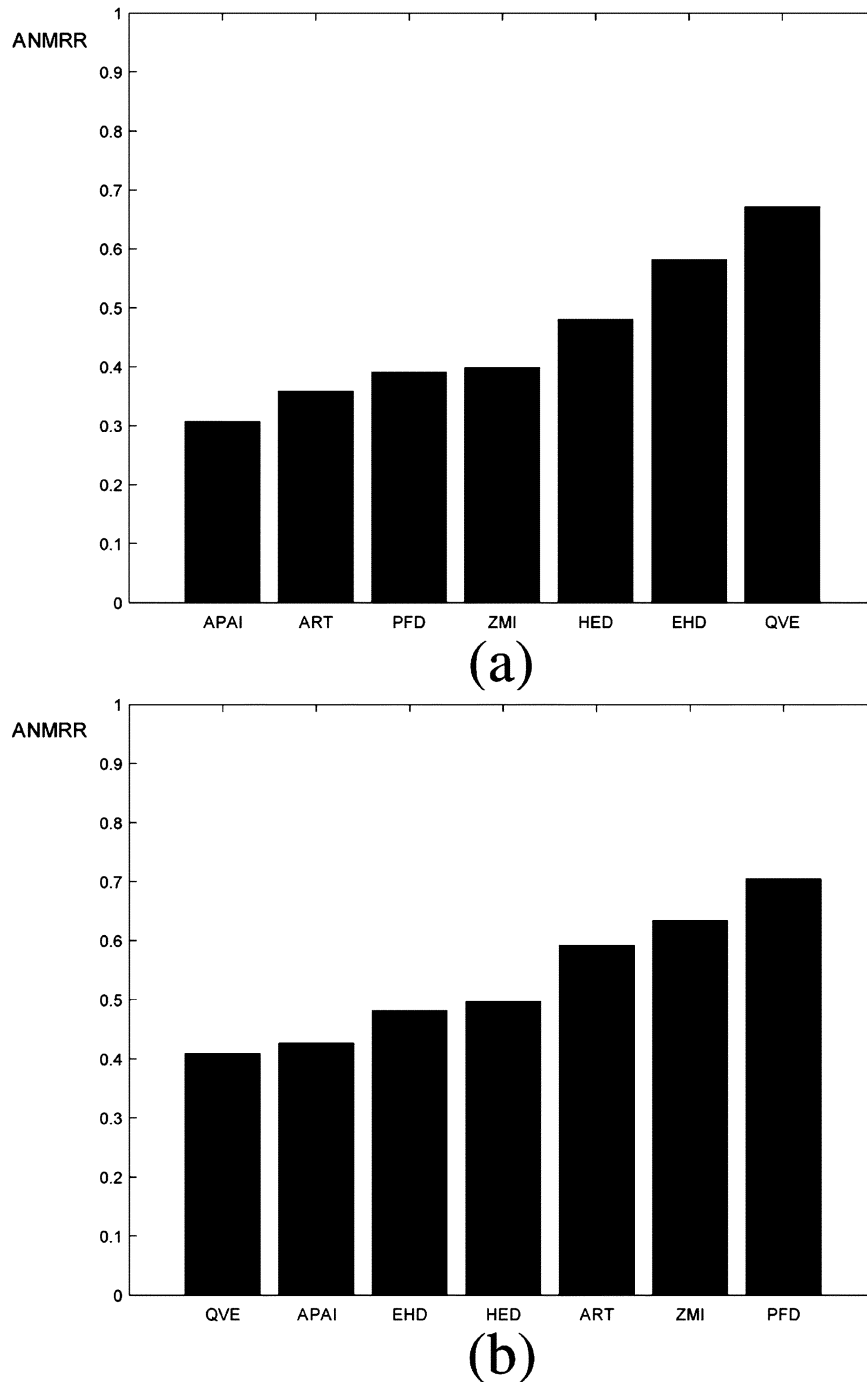


Fig. 6. ANMRR (average normalized modified retrieval rank) measure of different approaches based on (a) rotation invariance property and (b) scale invariant property. The abbreviations stand for the following approaches: angular partitioning of abstract image (APAI), angular radial transform (ART), polar Fourier descriptor (PFD), Zernike moment invariants (ZMI), histogram of edge directions (HED), MPEG-7's edge histogram descriptor (EHD), and query by visual example (QVE), respectively. The lower the ANMRR the better the performance.

### C. Comparative Results

To compare the retrieval performance of the proposed method with some other approaches within the literature, we implemented the following algorithms: 1) QVE as used in the QBIC system [3], [16]; 2) HED introduced by Jain and Vailaya [22]; 3) ZMI [24], [27]; 4) MPEG-7 EHD [28], [30]; 5) ART [9], [28]; and 6) polar Fourier descriptor (PFD) proposed by Zhang and Lu [15], [32]. The aim is to show the degree of rotation and scale invariance for these approaches. There are some other ap-

proaches for SBIR such as [10], introduced by Di Sciascio *et al.*, but since they need image segmentation at the preprocessing stage and the queries contain color and texture attributes, they are not applicable to this study.

It is also notable that the initial input to the ZMI, ART, and PFD methods is the thinned version of the sketched query and the strong edge map of the model image. This is to eliminate the adverse effect of color and texture diversity of the images on these methods. Moreover, in order to create a uniform assessment situation for all methods, we ignore the quantization

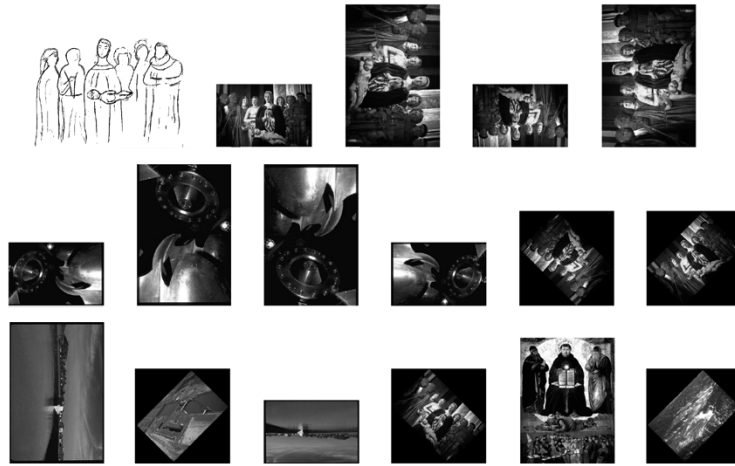


Fig. 7. Retrieval example 1: The top 16 retrieved images, in row order, when posing a sketched image (upper-left) with eight similar images in the ground truth,  $\text{NMRR} = 0.2400$ .

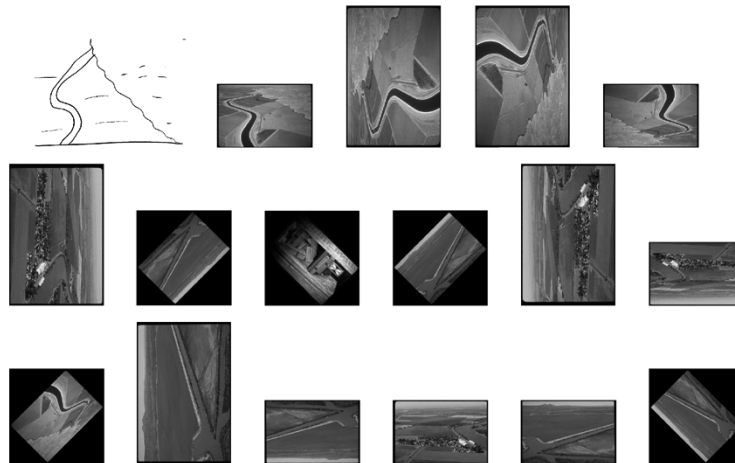


Fig. 8. Retrieval example 2: The top 16 retrieved images, in row order, when posing a sketched image (upper-left) with eight similar images in the ground truth,  $\text{NMRR} = 0.3600$ .

stage in the EHD and the ART methods. This will remove the retrieval performance disadvantage of quantization for the EHD and the ART methods.

All methods were tested using the ART-PHOTO BANK database. We applied  $k = 1$  in the HED method, resulting in a 70-entry feature vector. In the ART method, a 35-entry feature vector is achieved using  $m = 3$  and  $n = 12$ , as recommended in [28]. For the ZMI method, we used 36 moments as suggested in [24], resulting in a 36-entry feature vector. For the EHD method,  $\text{desired\_num\_of\_blocks}$  was set to 1100 and  $\text{Th}_{\text{edge}}$  was set to 11 (the default values) for the model images, and  $\text{Th}_{\text{edge}}$  was set to zero for the queries since they are binary images. A 150-bin histogram is obtained employing local, semiglobal and global histograms. Furthermore, we followed the algorithm given in [15] to obtain a 60-bin feature vector in the PFD approach. The proposed APAI method (Section III) resulted in a 70-bin feature vector using  $\beta = 3$ ,  $J = 257$ , and  $\varphi = 5^\circ$  as the internal values based on the discussion in the previous section.

The  $\ell_1$  (Manhattan) distance is used for measuring the similarity between all image features, while for the HED method a weighting factor of five for the global bins, as recommended in [30], was applied. The  $\ell_2$  (Euclidian) distance

was exploited for measuring the similarity between the PFD features [15], and a global correlation factor was employed for measuring the similarity between images in the QVE [16] method.

Fig. 6(a) shows the results expressed by the ANMRR measure. In our experiments, we used  $NG(q) = 8$  for all  $qs$ ,  $L = 16$  and  $Q_{\text{tot}} = 400$ . The results confirm that the proposed method yields the best retrieval performance (lowest ANMRR, i.e., 0.3070). The ART, PFD, and ZMI methods also show reasonable retrieval performance, i.e., 0.3589, 0.3908, and 0.3984, respectively. These methods perform better under the rotation test because their basis functions are designed specifically to be rotation invariant. The basis function used in the ART method [9] can efficiently capture the similarity among images as it splits the image (in the transform domain) into radial and angular directions.

The retrieval performance of the HED method is in a moderate level (0.4801), because the bins in the edge direction histogram are shifted during image rotation; therefore, rotation invariance is hardly achieved in this method. Although histogram smoothing is applied to overcome the problem [22], a better option might be a shift-invariant operator, such as absolute value

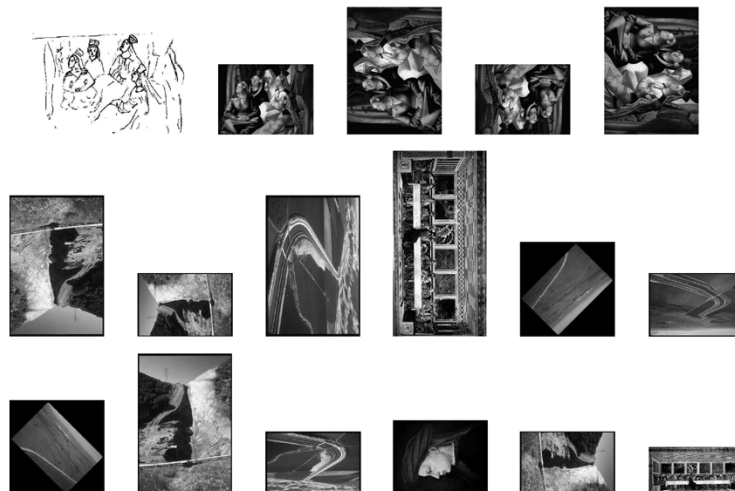


Fig. 9. Retrieval example 3: The top 16 retrieved images, in row order, when posing a sketched image (upper-left) with eight similar images in the ground truth,  $\text{NMRR} = 0.4200$ .

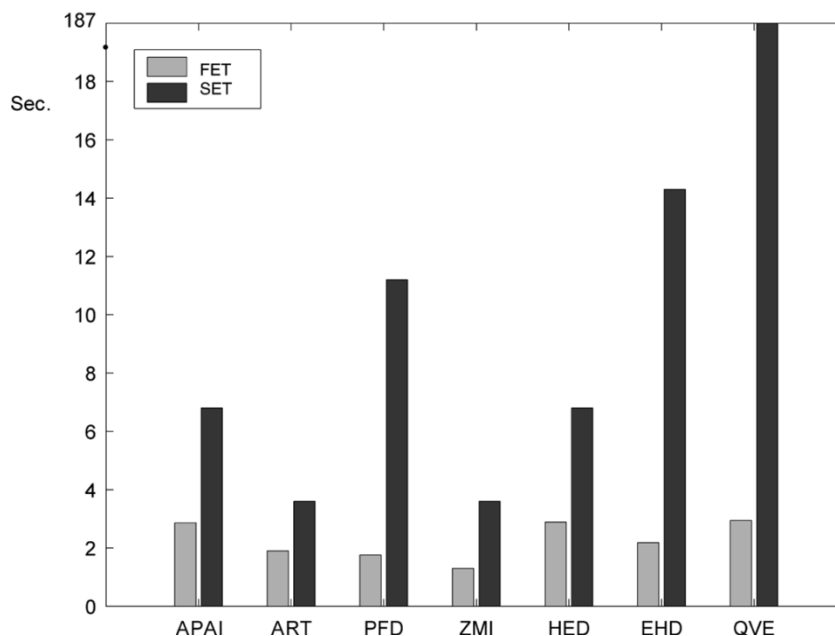


Fig. 10. FET and the SET for different methods. The vertical axis is discontinued, which shows that the search time for the QVE method is more than ten times longer than for the other methods.

of the Fourier transform to improve the rotation invariance property for this method.

The retrieval performances of the MPEG-7's EHD method (0.5816) and the QVE (0.6713) method unfold their lack of a rotation-invariance property. The EHD method considers only five predefined edge directions in local blocks which is inadequate to achieve rotation invariance. The QVE method compares neighboring pixels in the corresponding image regions and ignores global translation and rotation. Figs. 7–9 show three sets of retrieved images using the APAI method. For Fig. 7, the NMRR is 0.2400, and for Figs. 8 and 9, the NMRR is 0.3600 and 0.4200, respectively.

We also tested the scale invariance property of the methods in particular. All methods recruit size normalization, bounding box limitation, and/or center of mass alignment to achieve the scale and translation invariance properties. They exploit different nor-

malized sizes, for example, the QVE method normalizes images to  $64 \times 64$ , the ART method to  $101 \times 101$ , and the APAI method to  $257 \times 257$ . In this experiment, we applied the original 100 queries on a smaller database which contains 2000 full color images. This database includes the aforementioned 500 original images supplemented by scaled versions created by three scale factors of 0.5, 1.5, and 2. Once again, we obtained the ANMRR of the seven different approaches for these 100 queries with  $NG(q) = 4$  for all  $qs$ ,  $L = 8$ , and  $Q_{\text{tot}} = 100$ . Fig. 6(b) shows the results. As can be seen, the ANMRR measure of the proposed method, and the QVE, MPEG-7's EHD and HED methods are acceptable (less than 0.5). The QVE method shows the best retrieval performance in this test which confirms that the method well tolerates size variation and local translation. However, as already explored, the method is not rotation invariant and does not allow global rotations. It also needs more

computation power than the other methods since it calculates a global correlation factor between images, which significantly slows down the matching process. The resulting ANMRR of the other methods, i.e., ART, ZMI, and PFD, are high (more than 0.5). The reason seems to be that they concentrate on the rotation invariance property but are acutely sensitive to image size.

It is worthwhile to mention that all methods, except the QVE method, generate a feature vector for each image. Therefore, they can support indexing easily. The length of the feature vector (feature space dimension) for the selected methods are not exactly the same, i.e., 70, 36, 60, 36, 70, and 150 are the vector length for the APAI, ART, PDF, ZMI, EHD, and MPEG-7 EHD methods, respectively. On the other hand, the QVE approach, which uses a correlation scheme for measuring the similarity between images, cannot be used to generate indices for the database.

Finally, the feature extraction time (FET) and the SET, which are important factors for both the database population phase (off-line) and the query processing phase (on-line), were computed for the above methods. The average values of FET and SET, using the ART-PHOTO BANK database, are obtained using a PentiumIII, 1000-MHz machine and exhibited in Fig. 10. The FET parameters for the QVE, HED, APAI, and EHD methods are higher than for the ART, PFD, and ZMI methods. The reason being that the edge extraction, which is a time-consuming procedure, is only employed in the first group and not in the second group. The SET parameter, which deals directly with the comparison of feature vectors, shows proportionality to the corresponding vector's length with the exception of the PFD and QVE methods. The PFD method is the only one that uses the  $\ell_2$  distance for measuring the similarity between features, which needs more computation time than the  $\ell_1$  distance. The QVE method neglects feature vector advantages and calculates the global correlation between abstract images during the on-line phase. Therefore, its SET parameter is the highest (187 Sec.). Note the discontinuity in the time axis, which shows that the search time for the QVE method is more than ten times longer than for the other methods.

## V. CONCLUSION

The APAI approach presented in this paper enables measuring the similarity between a full color model image and a simple black and white sketched query. The images are arbitrary and may contain several complex objects in an inhomogeneous background. The approach deals directly with the entire image and needs no computationally intensive image segmentation and object extraction. Abstract images are defined based on the strong edges of the model image and the morphologically thinned outline of the query image. Angular partitioning of the abstract image, using the Fourier transform, is exploited to extract features that are scale and rotation invariant and robust against translation. Experimental results, using the APAI approach and the ART-PHOTO BANK as the test bed, show significant improvements in the ANMRR measure and rotation tolerance over six other well-known approaches within the literature. Individual tests on scale invariance were also

conducted and showed that the proposed method has better retrieval performance than five other approaches. While the QVE method exhibits the best tolerance for scale variations, its computational time is the highest (i.e., more than ten times longer). The proposed method depicts good retrieval performance in both rotation and scale tests, while it relies on reasonable feature extraction and SETs. The proposed intuitive user interface allows the user to interact with the system through a rough hand-drawn sketch without concern for precision, scale, orientation, or color. The pattern-matching capability of the system facilitates search and retrieval as well as other possible applications such as distortion measurement, drawing skill training and new generation system-man interfaces.

## ACKNOWLEDGMENT

The authors would like to acknowledge the World Art Kiosk, California State University, for providing paintings used in the ART-PHOTO BANK. They also thank Dr. N. Yasini, A. Mani, and K. Moosavian who kindly helped produce sketched queries.

## REFERENCES

- [1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1349–1380, Nov. 2000.
- [2] Y. Rui, T. Huang, and S. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Visual Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, 1999.
- [3] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC project: Querying images by content using color, texture, and shape," *Proc. SPIE*, vol. 1908, pp. 173–187, 1993.
- [4] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "PicSOM-content-based image retrieval with self-organizing maps," *Pattern Recognit. Lett.*, vol. 21, no. 13–14, pp. 1199–1207, 2000.
- [5] M. Beigi, A. B. Benitez, and S. Chang, "MetaSEEK: A content-based meta-search engine for images," *Proc. SPIE*, vol. 3312, pp. 118–128, 1997.
- [6] J. R. Smith and S. Chang, "VisualSEEK: A fully automated content-based image query system," in *Proc. ACM Multimedia*, 1996, pp. 87–98.
- [7] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.
- [8] B. S. Manjunath, J.-R. Ohm, and V. V. Vasudevan, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
- [9] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 716–719, Jun. 2001.
- [10] E. D. Sciascio, F. M. Donini, and M. Mongiello, "Spatial layout representation for query-by-sketch content-based image retrieval," *Pattern Recognit. Lett.*, vol. 23, no. 13, pp. 1599–1612, 2002.
- [11] H. H. S. Ip, A. K. Y. Cheng, W. Y. F. Wong, and J. Feng, "Affine-invariant sketch-based retrieval of images," in *Proc. IEEE Int. Conf. Comput. Graphics*, 2001, pp. 55–61.
- [12] A. Chalechale and A. Mertins, "An abstract image representation based on edge pixel neighborhood inform. (EPNI)," *Lecture Notes Comput. Sci.*, vol. 2510, pp. 67–74, 2002.
- [13] —, "Semantic evaluation and efficiency comparison of the edge pixel neighboring histogram in image retrieval," in *Proc. 1st Workshop Internet, Telecommun., Signal Process.*, 2002, pp. 179–184.
- [14] S. Matusiak, M. Daoudi, T. Blu, and O. Avaro, "Sketch-based images database retrieval," in *Proc. 4th Int. Workshop Adv. Multimedia Inform. Syst.*, 1998, pp. 185–191.
- [15] D. Zhang and G. Lu, "Shape-based image retrieval using generic fourier descriptor," *Signal Process.: Image Commun.*, vol. 17, no. 10, pp. 825–848, 2002.

- [16] K. Hirata and T. Kato, "Query by visual example-content based image retrieval," in *Proc. Adv. Database Technol.*, Berlin, Germany, 1992, pp. 56–71.
- [17] H.-W. Yoo, D.-S. Jang, S.-H. Jung, J.-H. Park, and K.-S. Song, "Visual information retrieval system via content-based approach," *Pattern Recognit.*, vol. 35, no. 3, pp. 749–769, 2002.
- [18] J. W. Lee, "A machine vision system for lane-departure detection," *Comput. Vision Image Understanding*, vol. 86, no. 1, pp. 52–78, 2002.
- [19] J.-L. Shih and L.-H. Chen, "A new system for trademark segmentation and retrieval," *Image Vision Comput.*, vol. 19, no. 13, pp. 1011–1018, 2001.
- [20] M. Abdel-Mottaleb, "Image retrieval based on edge representation," in *Proc. Int. Conf. Image Process.*, vol. 3, 2000, pp. 734–737.
- [21] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [22] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233–1244, Aug. 1996.
- [23] A. J. Jain and A. Vailaya, "Shape-based retrieval: A case study with trademark image databases," *Pattern Recognit.*, vol. 31, pp. 1369–1390, 1998.
- [24] "MPEG-7 Visual Part of Experimentation Model Version 5," Nordwijk-erhout, ISO/IEC JTC1/SC29/WG11/N3321, 2000.
- [25] C. Teh and R. T. Chin, "On image analysis by the method of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, Apr. 1988.
- [26] A. D. Bimbo, *Visual Inform. Retrieval*. San Francisco, CA: Morgan Kaufmann, 1999.
- [27] P. Sangassapaviriya, "Feature Extraction for Content-Based Image Retrieval," Master Thesis, Univ. Wollongong, NSW, Australia, 1999.
- [28] *MPEG-7 Visual Part of Experimentation Model Version 10.0*, Mar. 2001.
- [29] B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7 Multimedia Content Description Interface*. New York: Wiley, 2002.
- [30] C. S. Won, D. K. Park, and S. Park, "Efficient use of Mpeg-7 edge histogram descriptor," *Etri J.*, vol. 24, no. 1, pp. 23–30, Feb. 2002.
- [31] "A New Region-Based Shape Descriptor: The Art (Angular Radial Transform) Descriptor," ISO/IEC JTC1/SC29/WG11-MPEG99/M5472, 1999.
- [32] D. Zhang and G. Lu, "Enhanced generic fourier descriptors for object-based image retrieval," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 4, 2002, pp. 3668–3671.
- [33] —, "Generic fourier descriptor for shape-based image retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo.*, vol. 1, 2002, pp. 425–428.
- [34] L. Atzori and F. D. Natale, "Error concealment in video transmission over packet networks by a sketch-based approach," *Proc. Signal Proc.: Image Commun.*, vol. 15, no. 1–2, pp. 57–76, 1999.
- [35] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.
- [36] *Core Experiments on MPEG-7 Edge Histogram Descriptors*, May 2000.
- [37] R. O. Dida, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [38] J. P. M. D. Sa, *Pattern Recognition Concepts, Methods, and Applications*. New York: Springer-Verlag, 2001.



**Abdolah Chalechale** (S'03) received the B.S. degree in electrical engineering and M.Sc. degree in computer engineering from the Sharif University of Technology, Tehran, Iran, in 1991 and 1994, respectively. He is currently pursuing the Ph.D. degree at the University of Wollongong, NSW, Australia.

He was with Kermanshah Razi University from 1994 to 2001. His research interests include image and video retrieval, artificial intelligence in human-machine interface, and multimedia databases.



**Golshah Naghdy** was born in Tehran, Iran. She received the B.Sc. degree in electrical engineering and electronic engineering from Aryamehr University, Tehran, Iran, in 1977, the M.Phil. degree in control engineering from Bradford University, West Yorkshire, UK, in 1982, and the Ph.D. degree in electrical and electronic engineering from Portsmouth University, Hampshire, UK, in 1986.

She is an Associate Professor in the School of Electrical, Computer and Telecommunication Engineering, University of Wollongong, NSW, Australia.

She was a Senior Lecturer at Portsmouth University before joining Wollongong University in 1989. Her research interests include biological and machine vision in particular generic vision systems based on wavelet neurons. Another major area of her research is medical image processing. In this context, she has worked on three areas of medical image processing: loss-less coding of medical images for telemedicine or archiving, medical image enhancement, and mammogram image analysis for clinical decision support systems. She is currently working on the application of wavelet neurons in the development of artificial retina implants.



**Alfred Mertins** (M'94–SM'03) received the Dipl.-Ing. degree from the University of Paderborn, Paderborn, Germany, in 1984 and the Dr.-Ing. (Ph.D.) degree in electrical engineering and the Dr.-Ing. habil. degree in telecommunications from the Hamburg University of Technology, Hamburg, Germany, in 1991 and 1994, respectively.

From 1986 to 1991, he was with the Hamburg University of Technology, from 1991 to 1995 with the Microelectronics Applications Center, Hamburg, from 1996 to 1997 with the University of Kiel, Kiel, Germany, and from 1997 to 1998 with the University of Western Australia, Perth. From 1998 to 2003, he was with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, Australia. In April 2003, he joined the University of Oldenburg, Oldenburg, Germany, where he is a Professor with the Faculty of Mathematics and Science. His research interests include digital signal processing, digital communications, wavelets and filterbanks, and speech, audio, image, and video processing.