April 2004

Edge image description using angular radial partitioning

A. Chalechale
University of Wollongong

Alfred Mertins
University of Oldenburg, Germany, mertins@uow.edu.au

G. Naghdy
University of Wollongong, golshah@uow.edu.au

Edge image description using angular radial partitioning

Abstract

The authors present a novel approach for image representation based on geometric distribution of edge pixels. Object segmentation is not needed, therefore the input image may consist of several complex objects. For an efficient description of an arbitrary edge image, the edge map is divided into M/spl times/ N angular radial partitions and local features are extracted for these partitions. The entire image is then described as a set of spatially distributed invariant feature descriptors using the magnitude of the Fourier transform. The approach is scale- and rotation-invariant and tolerates small translations and erosions. The extracted features are characterised by their compactness and fast extraction/matching time. They exhibit significant improvement in retrieval performance using the average normalised modified retrieval rank (ANMRR) measure. Experimental results, using an image database initiated from a movie, confirm the supremacy of the proposed method.

Disciplines

Physical Sciences and Mathematics

Publication Details

Edge image description using angular radial partitioning

A. Chalechale, A. Mertins and G. Naghdy

Abstract: The authors present a novel approach for image representation based on geometric distribution of edge pixels. Object segmentation is not needed, therefore the input image may consist of several complex objects. For an efficient description of an arbitrary edge image, the edge map is divided into $M \times N$ angular radial partitions and local features are extracted for these partitions. The entire image is then described as a set of spatially distributed invariant feature descriptors using the magnitude of the Fourier transform. The approach is scale- and rotation-invariant and tolerates small translations and erosions. The extracted features are characterised by their compactness and fast extraction/matching time. They exhibit significant improvement in retrieval performance using the average normalised modified retrieval rank (ANMRR) measure. Experimental results, using an image database initiated from a movie, confirm the supremacy of the proposed method.

1 Introduction

Owing to an overwhelming increase in multimedia information in relevant databases, there is an urgent need for efficient tools to manage, search and retrieve such information. Multimedia storage and retrieval has been the focus of much research in recent years. The field also affects other disciplines, such as data compression, security and communication. MPEG-7 and CBIR (content-based image retrieval) are the two most important multimedia applications that have addressed this urgent need. MPEG-7 plans to provide a solution for the problem of efficient (fast) and effective (correct) retrieval through various multimedia materials. CBIR aims to facilitate the search in image databases based on the image content rather than text retrieval techniques.

In most current content-based image retrieval systems the emphasis is on four cues: colour, texture, shape and object layout. MPEG-7 suggests descriptors for colour, texture [1] and visual shape [2]. VisualSeek [3], also uses the object layout as an image feature. Although colour, texture and shape are significant features for retrieval purposes, they lose their original ability when the query or the database image has no such attributes, for example, when the query image is a rough sketch with only black and white lines [4], or when the aim is to search thousands of black and white trademarks without a well defined object contour for logos similar to a given one in a trademark registration process [5].

Rotation and translation invariant properties are crucial in most recognition tasks and should be considered in the features chosen for image retrieval. The invariant methods can be categorised into the following two main approaches:

• Image alignment, i.e. a transformation is applied to the image so that the object in the image is placed in a predefined standard position. Furthermore, the approach relies on the extraction of geometric primitives like extrema of the boundary curvature, bitangents or inflection points. Segmentation of the object is necessary and the approach is not trivial, especially when there exists more than one object in the image [6].
• Invariant features, i.e. using invariant image characteristics which remain unchanged if the object rotates or moves. Although this approach has attracted considerable interest [7], it is still based on geometric primitives. Therefore it suffers from the same shortenings as the image alignment approach.

It is desirable to avoid segmentation preprocessing and to start directly with the image pixels. One possibility is to employ invariant moments, such as regular moments or Zernike moments [8]. However, for eliminating image translations it is necessary to identify at least one matching point between images. The most common choice is the centre of mass for calculating central moments. Thus moments can be considered as a hybrid of alignment and invariant approaches.

The edge points hold considerable information about the image structure especially in the absence of colour/texture information or in images where colour and/or texture are not the discriminating factors. Furthermore, there are applications, such as sketch-based image retrieval, where only the edge map of the database image is comparable to the sketched query [4, 9, 10]. A face feature representation, called a line edge map (LEM), is proposed in [11] to integrate the structural information with spatial information of a face image by grouping pixels of face edge maps into line segments. In the edge pixel neighbourhood information (EPNI) method, neighbourhood structure of the edge pixels is used to make an extended feature vector [4]. The vector is used efficiently for comparing sketched queries with arbitrary images. The semantic power of the method is

A. Chalechale and G. Naghdy are with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia

A. Mertins is with the Signal Processing Group, Institute of Physics, University of Oldenburg, 26111 Oldenburg, Germany

examined in [12]. Although the method is scale- and translation-invariant, it does not exhibit rotation invariance.

Histograms of edge directions are widely used in computer vision and image retrieval for edge image matching. Jam and Vailaya propose an edge direction histogram for image retrieval [13] and employ it for trademark registration process. Shih and Chen [14] also use histograms of edge direction to describe the shapes of representative objects in different trademarks. Yoo *et al.* [15] apply the same histogram for shape representation in a new proposed content-based retrieval system. An edge distribution function is defined in [16]. A histogram of edge magnitude with respect to edge orientation angle is used for lane-departure detection. The edge histogram descriptor (EHD) is proposed in the MPEG-7 standard [17]. The descriptor can be modified for edge map matching by choosing internal parameters appropriately. The retrieval performance is improved by incorporating semi-global and global histograms to the traditional local histogram [18].

A set of angular radial transform (ART) coefficients is used as a region-based shape descriptor in [2, 17]. The descriptor describes the shape of an object in an image efficiently. The object may not only consist of a single closed and filled contour, but also objects with multiple contours are acceptable. Holes may exist inside the object in an image or in a frame of video sequence. This descriptor tolerates rotation and small translations.

Jia and Wang [19] have recently proposed a structural feature description based on geometric partitioning of edge images. They employ a sequential window sampling mode to partition edge pixels into circular blocks. The approach has a high computation cost. It also requires a predefined value (the number of edge pixels for each block) that need to be found empirically. The extracted feature vectors for different images have different sizes and the matching procedure is nontrivial.

In this paper, we develop a new approach for image matching using low-level features. It has to be emphasised that the images are arbitrary and may contain several complex objects. The method works directly with the edge points of the image so that object segmentation is not needed. It is based on accumulation of edge pixels in the image sectors defined by angular radial partitioning (ARP). The approach uses the magnitude of the Fourier transform in order to achieve rotation invariance. It is scale- and rotation-invariant and also tolerates small translations and erosions. The extracted feature vector is characterised by its compactness, fast extraction/matching time and significant retrieval performance. Its effectiveness and extraction time are compared with four other well known methods from the literature using the average normalised modified retrieval rank (ANMRR) measure.

2 Angular radial partitioning (ARP)

The main objective of ARP is to transform the image data into a new structure that supports measurement of the similarity between images in an effective, easy and efficient manner with emphasis on capturing scale and rotation invariant properties.

The edge map of an image carries the solid structure of the image, independent of the colour attribute. Its applicability is well known in computer vision, pattern recognition and image retrieval. Furthermore, in sketch-based image retrieval, it is the most useful feature to be employed for matching [4, 9, 10]. Therefore, at first the images in the database are converted to grey intensity by

eliminating the hue and saturation, while retaining the luminance. Applying an edge extraction operator, e.g. Canny edge operator [20], on this grey-scale image results in an edge image. In order to achieve the scale invariance property, the resulting edge image is then normalised to $W \times W$ pixels. This normalised edge image is called I and used for feature extraction. In the following, we consider pixels $I(\rho, \theta)$ to be either equal to '1' for edge pixels or '0' for non-edge pixels.

The algorithm uses the surrounding circle of I for partitioning it to $M \times N$ sectors, where M is the number of radial partitions and N is the number of angular partitions. The angle between adjacent angular partitions is $\theta = 2\pi/N$ and the radius of successive concentric circles is $\rho = R/M$, where R is the radius of the surrounding circle of the image (see Fig. 1).

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

$$f(k, i) = \sum_{\rho = \frac{kR}{M}}^{\frac{(k+1)R}{M}} \sum_{\theta = \frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}} I(\rho, \theta) \qquad (1)$$

for $k = 0, 1, 2 \ldots M - 1$ and $i = 0, 1, 2 \ldots N - 1$.

Although different lighting configurations yield different numbers of edge points in the sectors, this is not the case in hand-drawn sketches and also in many movie frames. To overcome the problem of different lighting conditions, the use of a curvature-insensitive edge detector is recommended. Detailed description of the best edge detector is beyond the scope of this work.

The feature extracted above will be circularly shifted when the image I is rotated $\tau = l2\pi/N$ rad $(l = 0, 1, 2 \ldots)$. To show this, let I_τ denote the image I after rotation by τ rad in a counterclockwise direction

$$I_\tau(\rho, \theta) = I(\rho, \theta - \tau) \qquad (2)$$

Then, it is obvious that

$$f_\tau(k, i) = f(k, i - l) \qquad (3)$$

where $f_\tau(k, i)$ are the image feature elements for I_τ for the same k and i, while $i - l$ is a modulo M subtraction. It means that there is a circular shift, for individual k s, in the image feature $\{f_\tau(k, i)\}$ representing I_τ, in comparison with the image feature $\{f(k, i)\}$ representing I.
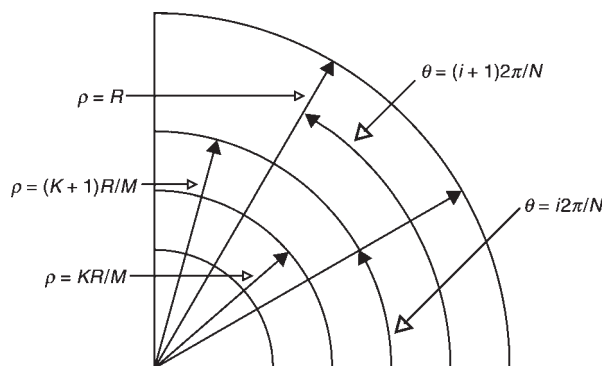

Fig. 1 *Angular radial partitioning of an image to N angular and M radial sectors where $k = 0, 1, 2 \ldots M - 1$ and $i = 0, 1, 2 \ldots N - 1$*
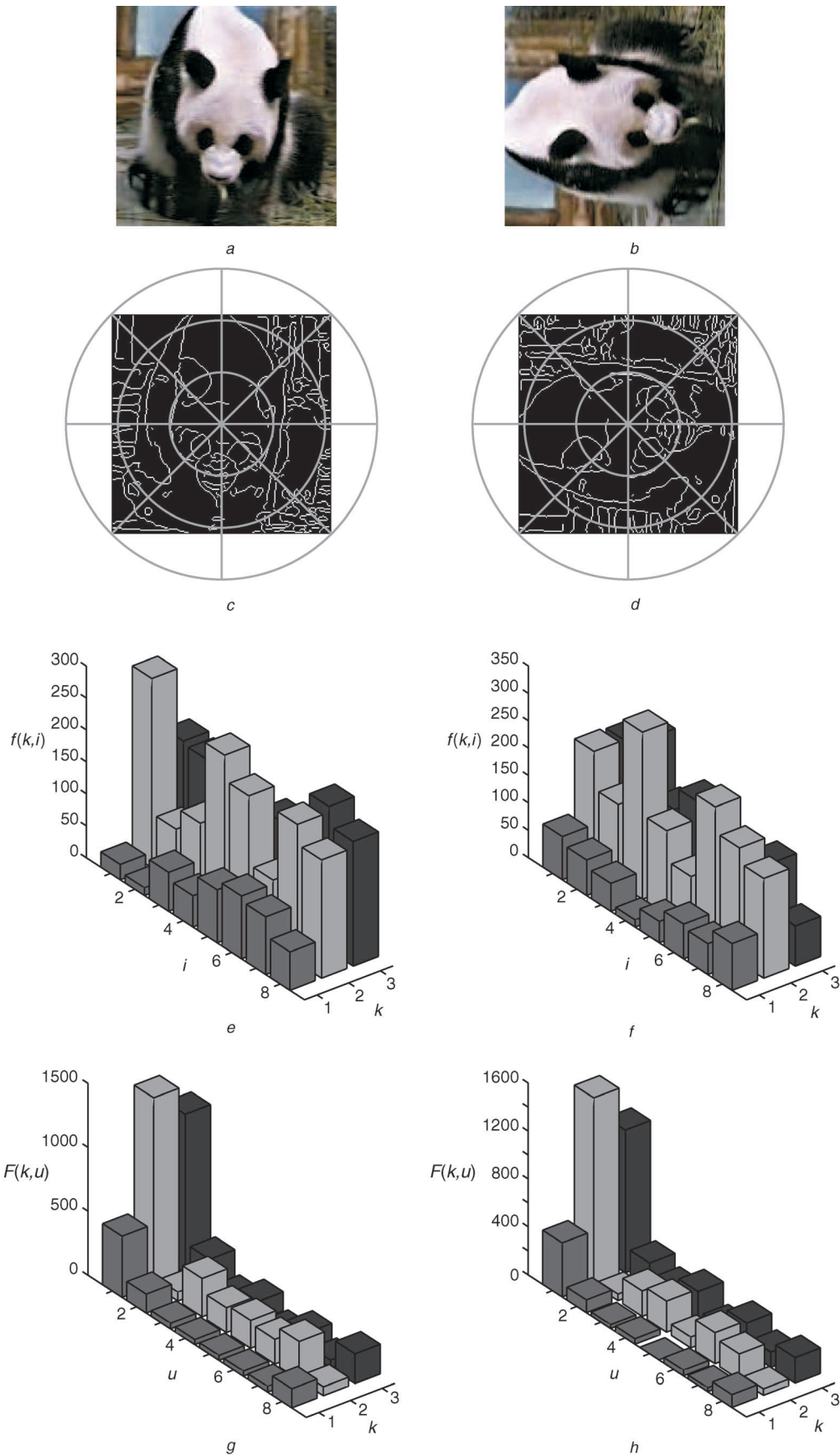
Fig. 2 *ARP example*

a Image example
b Its 90° rotated version
c, d Corresponding edge images superimposed with angular-radial partitions
e, f Number of pixels in the individual partitions
g, h Invariant features extracted using Fourier transform

Using a 1-D discrete Fourier transform of $f(k, i)$ and $f_\tau(k, i)$ for each k we obtain

$$F(k, u) = \frac{1}{N} \sum_{i=0}^{N-1} f(k, i) e^{-j 2\pi ui/N}$$

$$F_\tau(k, u) = \frac{1}{N} \sum_{i=0}^{N-1} f_\tau(k, i) e^{-j 2\pi ui/N}$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} f(k, i - l) e^{-j 2\pi ui/N}$$

$$= \frac{1}{N} \sum_{i=-l}^{N-1-l} f(k, i) e^{-j 2\pi u(i+l)/N}$$

$$= e^{-j 2\pi ul/N} F(k, u) \tag{4}$$

Because of the property $\|F(k, u)\| = \|F_\tau(k, u)\|$, the scale- and rotation-invariant image features are chosen as $\{\|F(k, u)\|\}$ for $k = 0, 1, 2 \ldots M-1$ and $u = 0, 1, 2 \ldots N-1$. Similarity between images is measured using ℓ_1 (Manhattan) distance between corresponding extracted features.

Choosing a medium-size sector (e.g. $M = 3$ and $N = 8$) makes the invariant image features extracted above robust to other small variations as well (i.e. translation, erosion and occlusion). This is due to the fact that the number of edge pixels in such sectors varies slowly with such variations.

Figure 2 shows an image example, its 90° rotated version, corresponding edge maps superimposed with ARP sectors, and extracted $f(k, i)$ and $F(k, u)$ features.

Experimental results (Section 4) show the robustness and retrieval effectiveness of the proposed method.

3 Alternative approaches

In this Section some other approaches that can be adapted for edge image matching are described.

3.1 Moment invariants

Moment invariants are widely used in pattern recognition and image analysis [5, 8, 21]. Geometric moments are defined as

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx\, dy \tag{5}$$

where $p, q = 0, 1, 2, \ldots, \infty$ and $f(x, y)$ is the density distribution function of the image. Seven invariant moment functions $(\phi_1 - \phi_7)$ are defined based on normalised central geometric moments (see [22]). The first six functions $(\phi_1 - \phi_6)$ are invariant under rotation, and the last one ϕ_7 is both skew and rotation invariant.

Zernike moments are less sensitive to noise and are more powerful in discriminating objects with n-fold symmetries than the geometric moments [23]. They are used for building a region-based shape descriptor in MPEG-7. Zernike orthogonal polynomials are employed to derive Zernike moment invariants of image $f(x, y)$ as follows:

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^{\infty} [V_{nl}(r, \theta)]^* f(r \cos\theta, r \sin\theta) r\, dr\, d\theta \tag{6}$$

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

$$V_{nl}(x, y) = R_{nl}(r) e^{il\theta} \tag{7}$$

and the radial polynomial is

$$R_{nl}(r) = \sum_{s=0}^{(n-|l|)/2} (-1)^s \cdot \frac{(n-s)!}{s!((n+|l|)/2 - s)!((n-|l|)2 - s)!} r^{n-2s}$$

$$= \sum_{\substack{k=|l| \\ n-k \text{ even}}}^{n} B_{n|l|k} r^k \tag{8}$$

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

$$f(x, y) \approx \sum_{n=0}^{N} \sum_{\substack{l \\ n-|l| \text{ even}, |l| \leq n}} A_{nl} V_{nl}(x, y) \tag{9}$$

The magnitudes of $\{A_{nl}\}$ are used for image matching and the similarity between images is measured using the ℓ_1 distance [21, 23]. In the case of edge images, $f(x, y)$ is a binary image representing the edge points. One obvious advantage of the use of binary functions is the low computational complexity. As $f(x, y)$ is an edge map, computation is not required for every pixel. In fact, computation is dependent on the existence of edge pixels.

3.2 Edge histogram descriptor (EHD)

The MPEG-7 standard defines the edge histogram descriptor (EHD) in its texture part [17]. 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. A given image is first divided into 16 sub-images (4×4), and local edge histograms are computed for each sub-image. To compute the edge histogram, each of the 16 sub-images 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:

$$x = \sqrt{\frac{image\_width * image\_height}{desired\_num\_of\_blocks}} \tag{10}$$

$$block\_size = \left\lfloor \frac{x}{2} \right\rfloor * 2 \tag{11}$$

where *image_width* and *image_height* represent the horizontal and vertical size of the image, respectively. Each image block is then partitioned into four (2×2) 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 edge images, the luminance takes only the value of one or zero. Edges are grouped into five classes: vertical, horizontal, 45° diagonal, 135° diagonal and isotropic (non-directional) based on directional edge strengths. These directions are determined for each image block using five corresponding 2×2 filter masks corresponding to 2×2 sub-divisions 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 grey-scale images is 11 and for binary edge images we set it to zero.

The histogram for each sub-image represents the frequency of occurrence of the five classes of edges in the corresponding sub-image. As there are 16 sub-images and each has a five-bin histogram, a total of $16 \times 5 = 80$ bins in the histogram is achieved. For normalisation, the number of edge occurrences for each bin is divided by the total number

96

of image blocks in the sub-image. To minimise the overall number of bits, the normalised bins are nonlinearly quantised and fixed-length coded with 3 bits per bin, resulting in a descriptor of size 240 bits.

Won *et al.* [18] proposed the efficient use of this descriptor by extending the histogram to 150 bins. The extended histogram is obtained by grouping the image blocks in 13 clusters (four vertical, four horizontal and five square clusters). Each cluster contains four sub-images. In addition to this semi-global histogram with $13 \times 5 = 65$ bins, another five-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 semi-global histograms could be produced directly from the local histogram at the matching time.

Although the approach achieves a good retrieval performance in some applications [17, 18], experimental results (Section 4) show that it does not exhibit the rotation invariance property since it considers a few predefined edge directions.

3.3 Histogram of edge directions (HED)

Histogram of edge directions (HED) for representing image information is one of the well known methods in the image retrieval literature [14–16]. Abdel-Mottaleb [24] utilises this method by applying the Canny edge operator to find strong edges in an image, and then quantises 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 [13] use edge directions as an image attribute for shape description. They show that, in the absence of colour information or in images with similar colours, this histogram is a significant tool in searching for similar images. They also exploit the histogram together with invariant moments in a case study using a trademark image database [5]. The edge information contained in the database images is extracted offline using the Canny edge operator. The corresponding edge directions are, subsequently, quantised into 72 bins of $5°$ each. To reduce the effect of rotation, they smooth the histogram as follows:

$$I_s[i] = \frac{\sum_{j=i-k}^{i+k} I[j]}{2k + 1} \qquad (12)$$

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

3.4 Angular radial transformation (ART)

The angular radial transform based shape descriptor is an efficient tool to retrieve object information. An ART-based descriptor is also adopted by MPEG-7 [17]. 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 disc that consists of the complete orthonormal sinusoidal basis functions in polar co-ordinates. From each image, a set of ART coefficients ψ_{mn} of order m and n is extracted using the following formula:

$$\psi_{mn} = \int_0^{2\pi} \int_0^1 V_{mn}^*(\rho, \theta) f(\rho, \theta) \rho \, d\rho \, d\theta \qquad (13)$$

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

$$V_{mn}(\rho, \theta) = A_n(\theta) R_m(\rho) \qquad (14)$$

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

$$A_n(\theta) = \frac{1}{2\pi} e^{jn\theta} \qquad (15)$$

and the radial basis function is defined by a cosine function

$$R_m(\rho) = \begin{cases} 1 & m = 0 \\ 2\cos(\pi m \rho) & m \neq 0 \end{cases} \qquad (16)$$

It can be shown that the magnitudes of the ART are rotation invariant [25]. The discrete ART coefficients of a grey-scale image can be found easily using a look-up table. Size normalisation, ART transformation and area normalisation are applied consecutively. In the first step, the size of the image is normalised by linear interpolation to a predefined width W and height H. Then the edge extraction process is carried out (e.g. by the Canny edge operator) to obtain the size-invariant edge map $I(x, y)$. The ART transformation is then applied. Finally, dividing the magnitude of each ART coefficient by the magnitude of the first coefficient (i.e. $\|\psi_{00}\|$) yields normalised ART coefficients that will be used for matching.

4 Experimental results

In order to evaluate retrieval effectiveness of the proposed method, first, we explain criteria used for measuring retrieval performance. Secondly, the effect of different parameters of the APP method are investigated and finally comparative results are presented.

4.1 Retrieval performance measure

Average normalised modified retrieval rank (ANMRR) is used as a retrieval performance measure. The ANMRR considers not only the recall and precision information, but also the rank information among the retrieved images. It is defined in MPEG-7 [26] as follows:

$$AVR(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} \qquad (17)$$

$$MRR(q) = AVR(q) - 0.5 - \frac{NG(q)}{2} \qquad (18)$$

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 * NG(q)} \qquad (19)$$

$$ANMRR = \frac{1}{Q} \sum_{q=1}^{Q} NMRR(q) \qquad (20)$$

where $NG(q)$ is the number of ground truth images for a query q. $K = \min(4 * NG(q), 2 * GTM)$, where GTM is $\max\{NG(q)\}$ for all q s of a data set. $Rank(k)$ is the rank of the found ground truth images, counting the rank of the first retrieved image as one. A *Rank* of $K + 1$ is assigned to

each of the ground truth images which are not in the first K retrievals. For example, suppose a given query q_i has ten similar images in an image database ($NG = 10$). If an algorithm finds six of them in the top 20 retrievals ($K = 20$) in the ranks of 1, 5, 8, 13, 14 and 18, then the $AVR(q_i) = 14.3$, $MRR(q_i) = 8.8$ and $NMRR(q_i) = 0.5677$.

Note that NMRR and its average (ANMRR) will always be in the range of [0,1]. Based on the definition of ANMRR, the smaller the ANMRR, the better the retrieval performance.

4.2 Effect of parameter variations in the ARP method

To show the effect of parameter variations on retrieval performance, different numbers of angular and radial partitions (N and M), using several size normalisation parameters W, are tested. We applied the ARP method on a database of 4320 images. The database was made by choosing 60 different pictures of the 'Animals have young'

movie from the MPEG-7 content set V14, and rotating each picture successively. The primary frame size was 352×288 pixels and each frame was rotated 72 times in $5°$ steps. To discard empty parts, created after rotation, we cropped the central 200×200 square and put it in our database. The cropped images are not only rotated versions of the original image, but also are slightly translated, with some extra and truncated parts near the borders (see Fig. 3 for some examples). The Canny operator, as an edge detector, with $\sigma = 1$ and Gaussian mask of size 9, is used to obtain the edge map of all images. Size normalisation, using nearest neighbour interpolation, is applied on the edge images. To evaluate the accuracy of the proposed method, we apply the original images as queries (60 images) while considering the rotated-cropped ones as database entries (4320 images).

Table 1 shows retrieval performance using three radial partitions with varying, coarse to fine, angular partitions. The observation indicates that, for all normalised sizes, the performance improved by increasing the number of angular partitions in most cases. There are, however, some
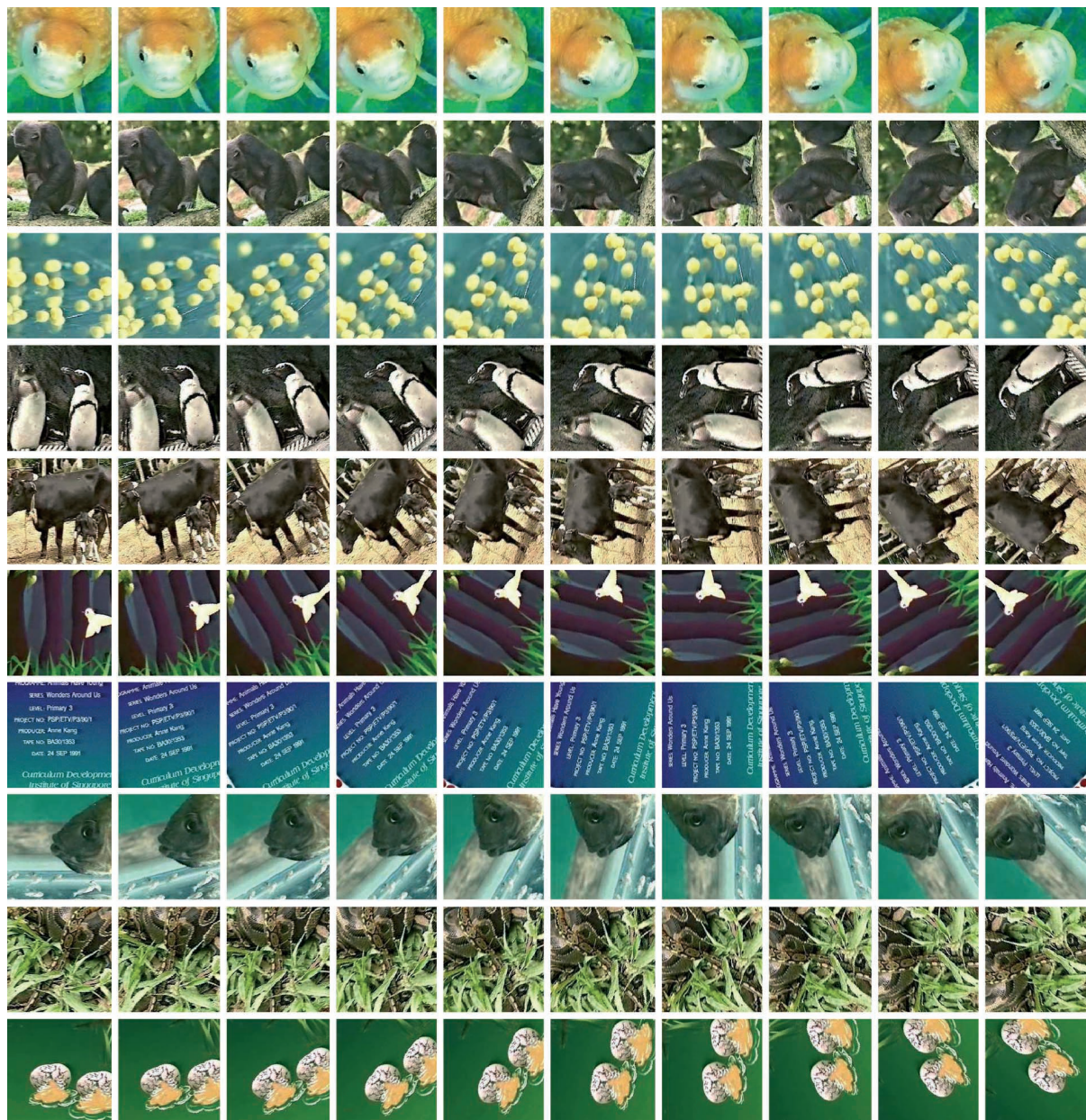

Fig. 3 *Image examples*
The first column shows the original images and the other columns are the variants (rotated, slightly translated and eroded)

Table 1: ANMRR of ARP method with three radial and varying angular partitions using main database (4320 images)

Normalised size	3×4 partitions	3×6 partitions	3×8 partitions	3×9 partitions	3×12 partitions	3×24 partitions	3×36 partitions	3×72 partitions
101×101	0.2921	0.2269	0.1538	0.1853	0.1922	0.1660	0.1661	0.1679
129×129	0.2625	0.1847	0.1090	0.1402	0.1441	0.0934	0.1010	0.0865
201×201	0.2072	0.1250	0.0552	0.0806	0.0832	0.0464	0.0451	0.0257
257×257	0.2216	0.1334	0.0694	0.0903	0.0931	0.0538	0.0518	0.0293

Table 2: ANMRR of ARP method with 12 angular and varying radial partitions using main database (4320 images)

Normalised size	5×12 partitions	7×12 partitions	10×12 partitions	15×12 partitions	18×12 partitions
101×101	0.1263	0.1107	0.1002	0.0952	0.0910
129×129	0.0974	0.0786	0.0716	0.0646	0.0566
201×201	0.0443	0.0317	0.0244	0.0217	0.0200
257×257	0.0512	0.0361	0.0269	0.0253	0.0223

exceptions in the Table, especially around $40°$ angular partitions. It appears that there are two contradictory factors: (*a*) benefits of capturing more details using small partitions, and (*b*) robustness against extra and truncated regions using large partitions. The exceptions in the results seem to be due to the advantage of medium size partitions which can tolerate small variations (rather than rotation) caused by translation and normalisation.

Table 2 exhibits ANMRR using 12 angular partitions with different radials. It is observed that increasing the number of radial partitions improves retrieval performance for all normalised sizes, with no exceptions.

Increasing the number of angular or radial partitions yields a bigger number of entries in the final feature vectors, which slows the online matching process. It also needs more storage and memory space. Therefore, a compromised decision on the number of partitions must be reached based on application in use.

The effect of normalisation is also depicted in Tables 1 and 2. The retrieval performance is the best in the normalised size of 201×201 as it is chosen close to the original image size. However, other sizes (101×101, 129×129 and 257×257) exhibit slow degradation of retrieval performance compared to 201×201 normalised size. It proves the robustness of the ARP method against varying image size.

In addition, to show rotation dependency of the proposed ARP method to the rotation step, we obtain retrieval performance for four other data sets. These include data sets of rotated images with the steps of $10°$, $15°$, $30°$ and $45°$. The extracted data sets contain 2160, 1440, 720 and 480 images, respectively. Again, we apply 60 different original images as queries while regarding rotated images as database entries. The results, as expressed in Table 3, indicate that using 3×12 partitions, the best performance is obtained on the image data set of the $30°$ step. However,

Table 3: ANMRR of the ARP method using selected data sets (2160, 1440, 720 and 480 images)

Partitions	10° rotated	15° rotated	30° rotated	45° rotated
3×12	0.0764	0.0789	0.0617	0.0972
3×8	0.0508	0.0557	0.0579	0.0472

using 3×8 partitions, the retrieval performance related to the data set of the $45°$ step is the best. The fact is based on matching of the rotated image with the angular partitions. However, the retrieval performance of other rotations is still reasonable.

The results presented exhibit the average performance. The retrieval process tolerates variations much better for some queries than others. The ℓ_1 (Manhattan) distance between two query examples (q_{10}, q_{52}) and the main database images (4320 images) are shown in Fig. 4. Each of 72 adjacent values are related to different rotated versions of each of 60 query images. q_{52} is an example where the ARP method tolerates variations as the ℓ_1 distances to the query's variations are less than the others (Fig. 4b). However, q_{10} is an example where the proposed method does not tolerate variations (rotation, translation, erosion and occlusion). The ℓ_1 distances of some other images are less than distance to the query's variations (Fig. 4a).

4.3 Comparative results

The Zernike moment invariants, edge histogram descriptor (EHD), histogram of edge directions (HED) and angular radial transformation (ART) methods are adapted and applied on the same test data. For Zernike moment invariants we used 36 moments as suggested in [21] resulting in a 36-entry feature vector. For the EHD method, the *desired_num_of_blocks* is set to 1100 (the default value) and (Th_{edge}) set to zero (because the edge images are all binary). A 150-bin histogram is obtained employing local, semi-global and global histograms. Furthermore, we use $k = 1$ in the HED method, resulting in a 70-entry feature vector, and a 35-entry feature vector is achieved using $m = 3$ and $n = 12$ in the ART method as recommended in [17]. Similarly, we set $M = 3$ and $N = 12$ in the proposed ARP method (Section 2) to make a same length (36-entry) feature vector.

In order to create a uniform assessment situation for all methods, we ignore the quantisation stage in the EHD and in the ART methods. This will remove the retrieval performance disadvantage of quantisation for the EHD and the ART methods. The ℓ_1 distance is used for measuring the similarity between all image features, while for the HED method a weighting factor of five for global bins, as recommended in [18], is applied.
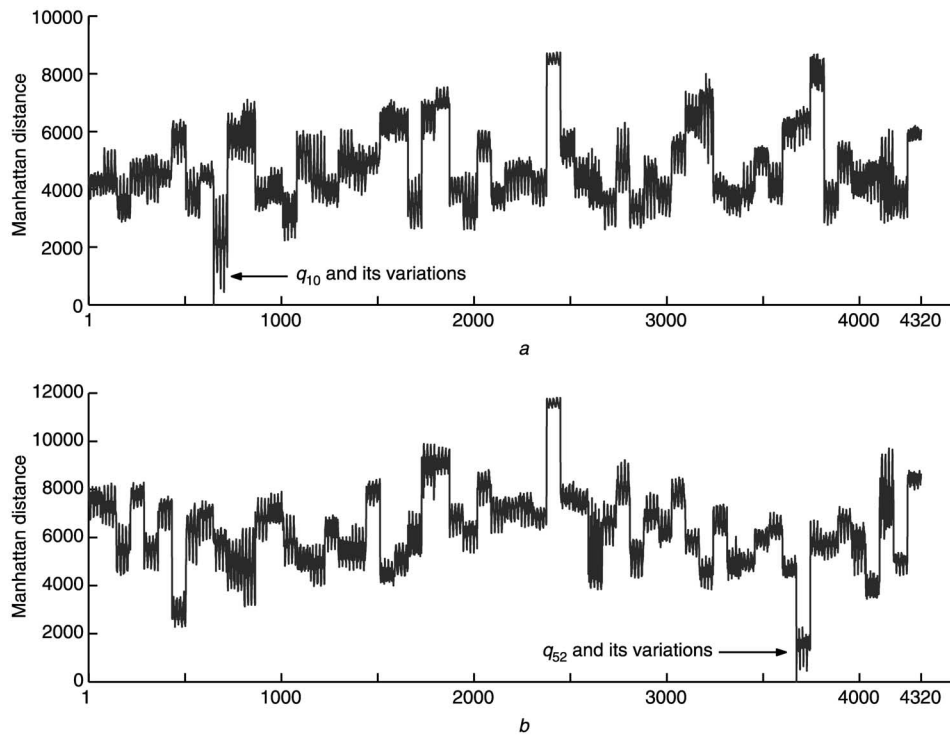
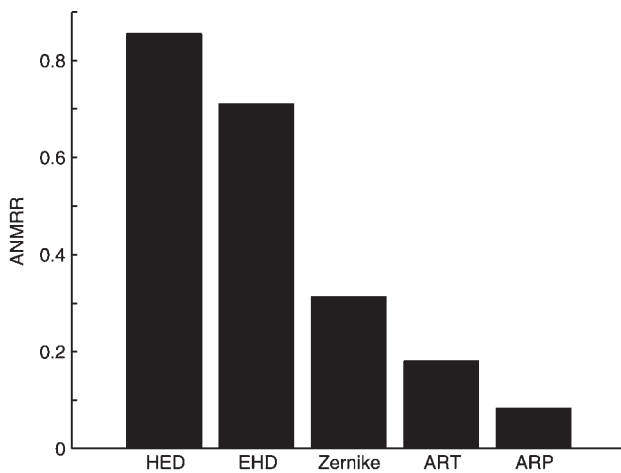**Fig. 4** *Manhattan (ℓ_1) distance of two query examples to the database images*

Fig. 5 *Retrieval results of different methods with ANMRR*

Fig. 6 *Comparison of feature extraction time*

Figure 5 shows the results expressed by the ANMRR. The results confirm that the proposed method yields the best retrieval performance (lowest ANMRR, i.e. 0.0832). The ART and the Zernike moment invariants methods also show reasonable retrieval performance, i.e. 0.1803 and 0.3127, respectively, but EHD (0.7106) and HED (0.8552) are not as robust to rotation as the others. It is also remarkable that the length of the feature vectors for the ARP, ART and Zernike moments methods are almost the same (36, 35, 36) while the length of the feature vector is 70 for the HED and 150 for the EHD methods.

Another comparison criterion, the feature extraction time, which is an important factor for online matching, was computed for the above methods. The average extraction time for one 200×200 edge image was computed as T, $5.8T$, $8.5T$, $12T$ and $149T$ for the ARP, EHD, HED, Zernike moments and ART methods, respectively, (see Fig. 6). The time T turned out to be 0.09 s using a Pentium-III, 1000 MHz machine. The proposed ARP method yields the shortest extraction time. The extraction time of the ART
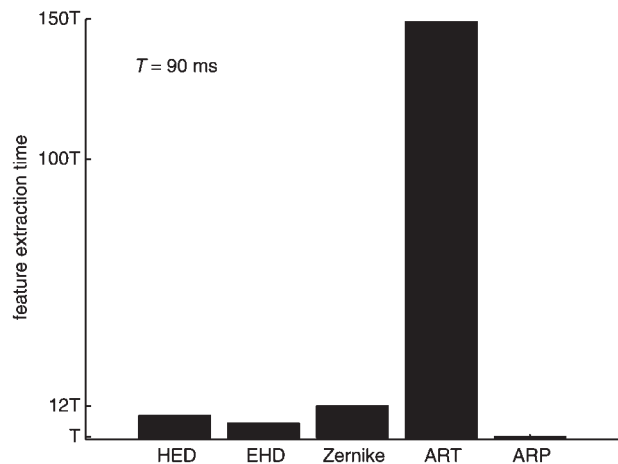
approach is too long, while the other methods exhibit acceptable extraction times.

5 Conclusions

We have introduced a novel approach for image representation based on geometrical distribution of edge pixels. Edge detection and size normalising are applied in the pre-processing stage and the resulting normalised edge map is the input of the main process. Object segmentation is not needed, therefore the input image may consist of several complex objects. For the efficient description of an arbitrary edge image, we propose dividing the edge image into M radial and N angular geometric partitions (sectors). The local features are computed with accumulating normalised edge pixels in the image sectors. The image is then represented as a set of spatially distributed feature descriptors. Applying the Fourier transform and using the magnitude of the transformed vectors ensures rotation invariance. The method is also scale invariant and tolerates

small translations and erosions. Image matching is achieved by measuring the distance between corresponding feature vectors. The compact feature vector (36 entries) not only accelerates the online matching process but also minimises the storage requirements. These characteristics make the approach attractive for large image database searching.

Experimental results, using an image database derived from a movie, show its supremacy in retrieval performance using the ANMRR factor as well as in the extraction time over four other methods adapted for edge image matching.

6 Acknowledgments

7 References

1 Manjunath, B.S., Ohm, J.-R., and Vasudevan, V.V.: 'Color and texture descriptors', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (6), pp. 703–715

2 Bober, M.: 'MPEG-7 visual shape descriptors', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (6), pp. 716–719

3 Smith, J.R., and Chang, S.: 'VisualSEEk: a fully automated content-based image query system'. Proc. ACM Multimedia 96, NY, USA, 1996, pp. 87–98

4 Chalechale, A., and Mertins, A.: 'An abstract image representation based on edge pixel neighborhood information (EPNI)', *Lect. Notes Comput. Sci.*, 2002, **2510**, (Proc. EurAsia-ICT 2002), pp. 67–74

5 Jain, A.J., and Vailaya, A.: 'Shape-based retrieval: a case study with trademark image databases', *Pattern Recognit.*, 1998, **31**, pp. 1369–1390

6 Huttenlocher, D.P., and Ullman, S.: 'Recognizing solid objects by alignment', *Int. J. Comput. Vis.*, 1990, **5**, (2), pp. 255–274

7 Weiss, I.: 'Geometric invariants and object recognition', *Int. J. Comput. Vis.*, 1993, **10**, (3), pp. 201–231

8 Teh, C.-H., and Chin, R.T.: 'On image analysis by the method of moments', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1988, **10**, (4), pp. 496–513

9 Ip, H.H.S., Cheng, A.K.Y., Wong, W.Y.F., and Feng, J.: 'Affine-invariant sketch-based retrieval of images'. Proc. IEEE Int. Conf. on Computer graphics, Hong Kong, 2001, pp. 55–61

10 Matusiak, S., Daoudi, M., Blu, T., and Avaro, O.: 'Sketch-based images database retrieval'. Proc. Advances in Multimedia Inform. Syst., 4th Int. Workshop, MIS'98, 1998

11 Gao, Y., and Leung, M.K.H.: 'Face recognition using line edge map', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (6), pp. 764–779

12 Chalechale, A., and Mertins, A.: 'Semantic evaluation and efficiency comparison of the edge pixel neighboring histogram in image retrieval'. Proc. WITSP'02 1st Workshop on the Internet, telecommunications and signal processing, Australia, 2002, pp. 179–184

13 Jain, A.K., and Vailaya, A.: 'Image retrieval using color and shape', *Pattern Recognit.*, 1996, **29**, (8), pp. 1233–1244

14 Shih, J.-L., and Chen, L.-H.: 'A new system for trademark segmentation and retrieval', *Image Vis. Comput.*, 2001, **19**, (13), pp. 1011–1018

15 Yoo, H.-W., Jang, D.-S., Jung, S.-H., Park, J.-H., and Song, K.-S.: 'Visual information retrieval system via content-based approach', *Pattern Recognit.*, 2002, **35**, (3), pp. 749–769

16 Lee, J.W.: 'A machine vision system for lane-departure detection', *Comput. Vis. Image Underst.*, 2002, **86**, (1), pp. 52–78

17 ISO/IEC JTC1/SC29/WG11/N4063, 'MPEG-7 Visual part of Experimentation Model Version 10.0', Singapore, March 2001

18 Won, C.S., Park, D.K., and Park, S.: 'Efficient use of MPEG-7 edge histogram descriptor', *ETRI J.*, 2002, **24**, (1), pp. 23–30

19 Jia, X.-M., and Wang, G.-Y.: 'Geometrical partition of edge image: a new approach for image structural features description'. Proc. 1st Int. Conf. on Machine learning and cybernetics, November 2002, pp. 1391–1395

20 Canny, J.: 'A computational approach to edge detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, November 1986, **8**, (6), pp. 679–698

21 ISO/IEC JTC1/SC29/WG11/N3321, 'MPEG-7 visual part of experimentation model version 5', Nordwijkerhout, March 2000

22 Bimbo, A.D.: 'Visual information retrieval' (Morgan Kaufmann Publishers, 1999)

23 Sangassapaviriya, P.: 'Feature extraction for content-based image retrieval'. Masters thesis, University of Wollongong, Australia, 1999

24 Abdel-Mottaleb, M.: 'Image retrieval based on edge representation'. Proc. Int. Conf. on Image processing, Piscatway, NJ, USA, 2000, vol. 3, pp. 734–737

25 ISO/IEC JTC1/SC29/WG11- MPEG99/M5472, 'A new region-based shape descriptor: The art (angular radial transform) descriptor'. Maui, December 1999

26 ISO/IEC JTC1/SC29/WG11- MPEG/M5984, 'Core experiments on MPEG-7 edge histogram descriptors', Geneva, Switzerland, May 2000