Angular Radial Partitioning for Edge Image Description

Abdolah Chalechale[1] and Alfred Mertins[2]

[1] School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW 2522, Australia
ac82@uow.edu.au

[2] Signal Processing Group, Institute of Physics
University of Oldenburg, 26111 Oldenburg, Germany
alfred.mertins@uni-oldenburg.de

Abstract

A novel approach for image representation based on geometric distribution of edge pixels is presented. Object segmentation is not needed, therefore, the input image may consist of several complex objects. The edge map is divided into $M \times N$ angular radial partitions and local features are extracted for these partitions. The approach is scale and rotation invariant and also tolerates small translations and erosions. The extracted features are characterized by their compactness and fast extraction/matching time. They exhibit significant improvement in retrieval performance using ANMRR measure. Experimental results show their supremacy using an image database initiated from a movie.

1 Introduction

Based on the ever increasing amount of multimedia information in relevant databases and also on the Web there is an urgent need for efficient tools to manage and search in such data. Multimedia storage and retrieval have attracted many new researches. They also affect other disciplines such as compression, security and communication. MPEG-7 and CBIR (Content-Based Image Retrieval) are the two most important multimedia applications that have addressed this urgent issue. MPEG-7 plans to provide a solution for the problem of efficient (fast) and effective (correct) retrieval through various types of multimedia material. CBIR's aim is 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 clues: color, texture, shape and the objects' layout. MPEG-7 suggests descriptors for color and texture [1] and for visual shape [2]. VisualSEEk [3] uses the object layout as an image feature. Although color, texture and shape are significant features for retrieval purposes, they lose their original importance when the query or the database image has no such attributes. This is for example the case when the query image is a fast drawn, rough sketch with only some black and white lines [4], or when the aim is to search in thousands of black and white trademarks without a well defined object contour [5].

Rotation and translation invariant properties are crucial in most recognition tasks and should be considered in the features applied in image retrieval. The invariant methods can be categorized in 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. The approach relies furthermore 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 scene [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 and is therefore suffering from the same difficulties as already mentioned for image alignment.

It is desirable to avoid preprocessing for segmentation 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 center of mass for calculating central moments. Thus moments can be considered as a hybrid of alignment and invariants.

The edge points hold considerable information about the image structure especially in the absence of color/texture information or in the presence of images with similar color/texture. 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 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. Edge Pixel Neighborhood Information (EPNI) is using neighborhood structure of the edge pixels 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 examined in [12]. Although the method is scale and translation invariant it does not exhibit the rotation invariance.

A histogram of edge directions was introduced by Jain and Vailya [13] and is widely used as an invariant image feature [5, 14]. In this method, the smoothed histogram of the edge point directions is used to find the similarity between images. Edge Histogram Descriptor (EHD) was proposed in the MPEG-7 standard [15]. 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 [16].

A set of Angular Radial Transform (ART) coefficients is used as region-based shape descriptor in [2, 15]. 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, and also objects with multiple contours are acceptable. Holes may exist inside the object in an image or frame of video sequence. This descriptor tolerates rotation and small translations.

Jia and Wang [17] recently proposed a structural feature description based on geometric partitioning of edge images. They used 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 present a new approach for image matching using low-level features. It has to be emphasized 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 accumulating 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 feature vector extracted is characterized by its small size, fast extraction/matching time, and significant retrieval performance. Its effectiveness and extraction time are compared with four other methods known from the literature using ANMRR measure.

The rest of the paper is organized as follows. Section 2 describes the proposed approach in details. Section 3 presents comparative results, and Section 4 concludes the paper.

2 Angular Radial Partitioning (ARP)

The algorithm for extracting ARP is given in this section. The main objective of the algorithm is to transform the image data into a new structure that supports measurement of the similarity between images in a correct, easy and fast way 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 color 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 that can be employed for matching [4, 9, 10]. Therefore, at first the color image is converted to a gray intensity image by eliminating the hue and saturation while retaining the luminance. The image is then normalized to 201×201 pixels. Applying the Canny edge operator [18] on this normalized gray scale image results in an edge image I, which is employed 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 $i = 0, 1, 2 \ldots N-1$. The feature extracted above will be circularly shifted when the image I is rotated $\tau = l2\pi/N$ radian ($l = 0, 1, 2 \ldots$). To show this, let I_τ denote the edge map I after rotation by τ radians in counterclockwise direction:

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

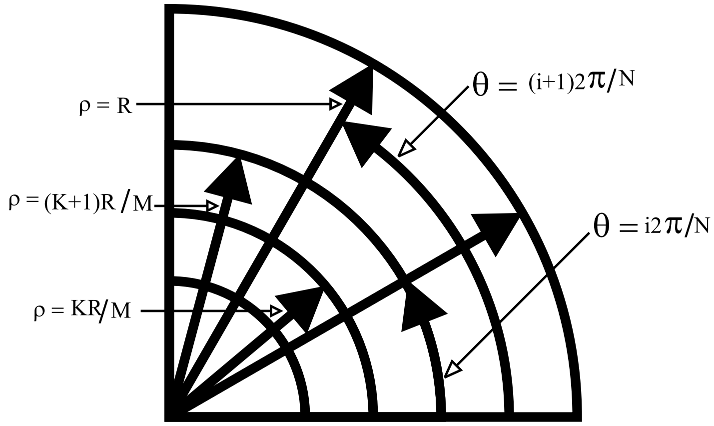

Figure 1: Angular Radial Partitioning of an image to N angular and M radial sectors where $k = 0, 1, 2 \ldots M - 1$ and $i = 0, 1, 2 \ldots N - 1$.

.

Then,

$$f_\tau(k, i) = \sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}} \sum_{\theta=\frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}} I_\tau(\rho, \theta) \tag{3}$$

are the image feature elements for I_τ for the same k and i. We can express f_τ as

$$\begin{aligned} f_\tau(k,i) &= \textstyle\sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}} \sum_{\theta=\frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}} I(\rho, \theta - \tau) \\ &= \textstyle\sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}} \sum_{\theta=\frac{(i-l)2\pi}{N}}^{\frac{(i-l+1)2\pi}{N}} I(\rho, \theta) \\ &= f(k, i-l) \end{aligned} \tag{4}$$

where $i - l$ is a modulo M subtraction. It means that there is a circular shift (for individual k's) in the image feature $\{f_\tau(k, i)\}$ regarding to the image feature $\{f(k, i)\}$ which representing I_τ and I respectively.

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

$$\begin{aligned} F(k,u) &= \textstyle\frac{1}{N}\sum_{i=0}^{N-1} f(k,i)e^{-j2\pi ui/N} \\ F_\tau(k,u) &= \textstyle\frac{1}{N}\sum_{i=0}^{N-1} f_\tau(k,i)e^{-j2\pi ui/N} \\ &= \textstyle\frac{1}{N}\sum_{i=0}^{N-1} f(k,i-l)e^{-j2\pi ui/N} \\ &= \textstyle\frac{1}{N}\sum_{i=-l}^{N-1-l} f(k,i)e^{-j2\pi u(i+l)/N} \end{aligned} \tag{5}$$

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 $u = 0, 1, 2 \ldots N - 1$.

Choosing a medium-size sector (e.g. $M = 3$ and $N = 12$) makes the invariant image feature extracted above to be robust to other small variations as well (i.e. translation, erosion and occlusion). This is based on the fact that the number of edge pixels in such sectors varies slowly with such variations.

Experimental results (Section 3) show its robustness and effective retrieval performance.

3 Experimental Results

Comparative results are presented in this section. We used a database of 4320 images. The database was made by choosing 60 different pictures of *Animals have young* movie from the MPEG-7 content set V14, and then each picture is rotated successively. The original frame size was 352×288 pixels and each frame was rotated 72 times in $5°$ steps. We cropped the central 201×201 square and put it in the database. The cropped image is not only a rotated version of the original image but also a little translated with some extra or truncated parts at the borders (see Figure 2 for some examples). Canny edge operator [18] was used to obtain the edge map of all images. To evaluate the accuracy of the proposed method, we applied original images as queries (60 images) while considering the rotated-cropped ones as image database (4320 images).

The Zernike moment invariants [19], edge histogram descriptor (EHD) [16], histogram of edge directions (HED) [13], and angular radial transformation (ART) [15] methods are also applied on the same test data. For Zernike moment invariants we used 36 moments as suggested in [19], resulting in a 36-entry feature vector. For EHD method, *desired_num_of_blocks* was set to 1100 (the default value) and (Th_{edge}) set to zero (because the edge images are all binary). A 150-bin histogram was obtained employing local, semi-global and global histograms. Furthermore, we used $k = 1$ in HED method, resulting in a 70-entry feature vector, and a 35-entry feature vector was achieved using $m = 3$

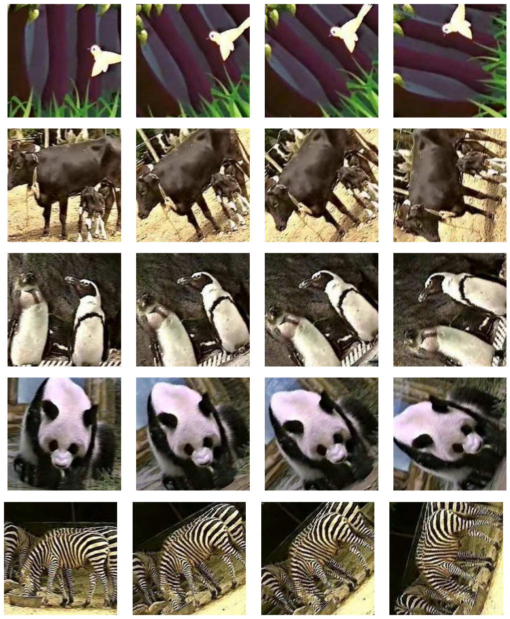

Figure 2: Image examples, the cropped-rotated versions have small translations while extra and eroded parts near the borders.

and $n = 12$ in ART method. Setting $M = 3$ and $N = 12$ in the proposed ARP method (Section 2) led to a 36-entry feature vector.

We ignored the quantization stage in EHD and ART methods to put all methods in the same situation. The ℓ_1 distance was used for measuring the similarity between all image features, while for the HED method a weighting factor of 5 for global bins, as recommended in [16], was applied.

We used Average Normalized Modified Retrieval Rank (ANMRR) [20] for measuring retrieval accuracy. Although Precision and Recall are well-known measures for the retrieval performance, they do not consider the rank of retrieval. ANMRR is a measure that exploits both the number of similar retrieved images and also their rank in the output list. ANMRR is defined as follows [21]:

$$AVR(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} \tag{6}$$

$$MRR(q) = AVR(q) - 0.5 - \frac{NG(q)}{2} \tag{7}$$

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 * NG(q)} \tag{8}$$

$$ANMRR = \frac{1}{Q} \sum_{q=1}^{Q} NMRR(q) \tag{9}$$

$NG(q)$ is the number of ground truth images for a query q and $Rank(k)$ is the rank of the retrieved image in the ground truth. $K = \min(4 * NG(q), 2 * GTM)$ where GTM is max $\{NG(q)\}$ for all q's of a data set. 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. In our experiments the $NG(q) = 72$ for all q's, $K = 144$ and $Q = 60$.

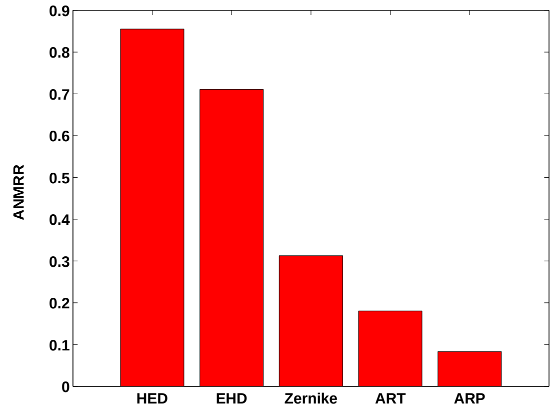

Figure 3: Retrieval results of different methods with ANMRR.

Figure 3 shows the results expressed by the ANMRR. As one can see in the figure, the proposed method yields the best retrieval performance (lowest ANMRR i.e. 0.0963). The ART and 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 feature vectors' length of ARP, ART and Zernike moments methods are almost the same (36,35,36) while the vector's length is 70 for HED and 150 for EHD methods.

As 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 201×201 edge image was computed as T, $5.8T$, $8.5T$, $12T$ and $149T$ for ARP, EHD, HED, Zernike moments and ART methods respectively (see Figure 4). The time T turned out to be 0.09 seconds using a Pentium-III, 1000 MHz machine. The proposed ARP method possessed the shortest extraction time. The extraction time of ART approach is too long while the other methods exhibit acceptable extraction times.

4 Conclusion

We have introduced a novel approach for image representation based on geometrical distribution of edge pixels. Edge detection is applied in the pre-processing stage and the resulting edge map is the input of the approach. Object segmentation is not needed, thereby the input image may consist of several complex objects. For efficient description of an arbitrary edge image, we proposed dividing the edge image into M radial and N angular geometric partitions (sectors). The local features are computed accumulating normalized edge pixels in the image sectors. The im-

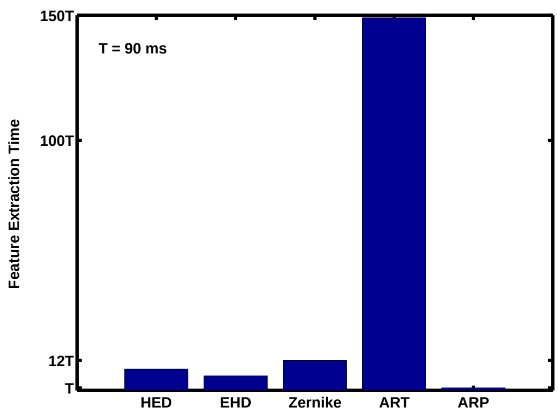

Figure 4: Comparison of feature extraction time.

age is then represented as a set of spatially distributed feature descriptors. Applying Fourier transform and using magnitude of the transformed vectors makes it rotation invariant. It is also scale invariant and tolerable to small translations and erosions. Image matching is achieved by measuring the distance between corresponding feature vectors. The features extracted form a 36-entry vector not only accelerate the online matching process but also minimize the storage requirements. These characteristics make the approach a suitable solution for large image database searching.

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

Acknowledgments. The first author is financially supported by the Ministry of Science, Research and Technology of I.R. Iran.

References

[1] B. S. Manjunath, J.-R. Ohm, and V. V. Vasudevan, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, June 2001.

[2] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716–719, June 2001.

[3] J. R. Smith and S. Chang, "VisualSEEk: a fully automated content-based image query system," in *Proceedings ACM Multimedia 96.*, NY, USA, 1996, pp. 87–98.

[4] A. Chalechale and A. Mertins, "An abstract image representation based on edge pixel neighborhood information (EPNI)," in *Lecture Notes in Computer Science, EurAsian-ICT 2002:information and communication technology*, 2002, vol. 2510, pp. 67–74.

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

[6] D. P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment," *International journal of computer vision*, vol. 5, no. 2, pp. 255–274, 1990.

[7] I. Weiss, "Geometric invariants and object recognition," *International journal of computer vision*, vol. 10, no. 3, pp. 201–231, 1993.

[8] C.-H. Teh and R. T. Chin, "On image analysis by the method of moments," *IEEE transactions on pattern analysis and machine intelligence*, vol. 10, no. 4, pp. 496–513, 1988.

[9] H. H. S. Ip, A. K. Y. Cheng, W. Y. F. Wong, and J. Feng, "Affine-invariant sketch-based retrieval of images," in *Proceedings. Computer Graphics International. IEEE Comput. Soc*, 2001, pp. 55–61.

[10] S. Matusiak, M. Daoudi, T. Blu, and O. Avaro, "Sketch-based images database retrieval," in *Advances in Multimedia Information Systems. 4th International Workshop, MIS'98. Proceedings.*, 1998.

[11] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764–779, 2002.

[12] A. Chalechale and A. Mertins, "Semantic evaluation and efficiency comparison of the edge pixel neighboring histogram in image retrieval," in *WITSP'02 first workshop on the internet, telecommunications and signal processing*, Australia, 2002, pp. 179–184.

[13] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, Aug. 1996.

[14] M. Abdel-Mottaleb, "Image retrieval based on edge representation," in *Proceedings 2000 International Conference on Image Processing*, Piscatway, NJ, USA, 2000, vol. 3, pp. 734–737.

[15] ISO/IEC JTC1/SC29/WG11/N4063, "MPEG-7 visual part of experimentation model version 10.0," Singapore, March 2001.

[16] C. S. Won, D. K. Park, and S. Park, "Efficient use of Mpeg-7 edge histogram descriptor," *Etri Journal*, vol. 24, no. 1, pp. 23–30, Feb. 2002.

[17] X.-M. Jia and G.-Y. Wang, "Geometrical partition of edge image: a new approach for image structural features description," in *Proceeding of the first International conference on machine learning and cybernetics*, November 2002, pp. 1391–1395.

[18] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[19] ISO/IEC JTC1/SC29/WG11/N3321, "MPEG-7 visual part of experimentation model version 5," Nordwijkerhout, March 2000.

[20] ISO/IEC JTC1/SC29/WG11/N2929, "Core Experiments for MPEG-7 Color/Texture Descriptors," Melbourn, Oct. 1999.

[21] ISO/IEC JTC1/SC29/WG11-MPEG2000/M5984, "Core experiments on MPEG-7 edge histogram descriptors," Geneva, May 2000.