Published in IET Image Processing
Received on 30th August 2009
Revised on 9th March 2010
doi: 10.1049/iet-ipr.2009.0246

Special Issue on Imaging Systems and Techniques

ISSN 1751-9659

Evaluation of shape descriptors for shape-based image retrieval

A. Amanatiadis[1] *V.G. Kaburlasos*[2] *A. Gasteratos*[1]
S.E. Papadakis[2]

[1]*Laboratory of Robotics and Automation, Department of Production and Management Engineering, Democritus University of Thrace, University Campus Kimmeria, GR-67100 Xanthi, Greece*
[2]*Division of Computing Systems, Department of Industrial Informatics, Technological Educational Institution of Kavala, GR-65404 Kavala, Greece*
E-mail: aamanat@ee.duth.gr

Abstract: This article presents a comparative study between scale, rotation and translation invariant descriptors for shape representation and retrieval. Since shape is one of the most widely used image feature exploited in content-based image retrieval systems, the authors studied for each descriptor, the number of coefficients needed for indexing and their retrieval performance. Specifically, the authors studied Fourier, curvature scale space, angular radial transform (ART) and image moment descriptors for shape representation. The four shape descriptors are evaluated against each other using the standard methodology and the two most appropriate and available databases. The results showed that moment descriptors present the best performance in terms of shape representation quality while ART presents the lowest descriptor size.

1 Introduction

Shape representation compared to other features, like texture and colour, is much more effective in semantically characterising the content of an image [1]. However, the challenging task of shape descriptors is the accurate extraction and representation of shape information. The construction of shape descriptors is even more complicated when invariance, with respect to a number of possible transformations, such as scaling, shifting and rotation, is required [2]. The overall performance of shape descriptors can be divided into qualitative and quantitative performances. The qualitative characteristics involve their retrieval performance based on the captured shape details for representation. Their quantitative performance includes the amount of data needed to be indexed in terms of number of descriptors, in order to meet certain qualitative standards [3] as well as their retrieval computational cost.

While studies have been extended to content-based three-dimensional (3D) shape retrieval methods [4], still pattern recognition by 2D shape descriptors can be used in many practical tasks, for example in image matching, multitemporal image sequence analysis, shape classification and character recognition. Furthermore, their quantitative characteristics which still remain superior, make them widely usable and effective [5].

Various shape descriptors exist in the literature, mainly categorised into two groups: contour-based shape descriptors and region-based shape descriptors. Contour-based methods need extraction of boundary information which in some cases may not available. Region-based methods, however, do not rely on shape boundary information, but they take into account all the pixels within the shape region. Therefore for generic purposes, both types of shape representations are necessary.

An extensive evaluation of MPEG-7 shape descriptors have been presented in [6] where the effectiveness of Zernike moments and Fourier descriptors (FD) was confirmed through experimental results. In [7], the curvature scale space descriptor outperforms the other evaluated shape descriptors when compared in the MPEG-7 Core Experiment. In [8], descriptors based on complex moments and spectral transforms, such as Zernike moments and FD, are proved to be the best choices for general shape applications. While the aforementioned descriptors are some of the most important shape descriptors, they have never been evaluated against each other.

In this paper, we review and evaluate these important shape descriptors using the MPEG-7 Core Experiment shape database. Their retrieval performance and comparison results are presented and discussed. Furthermore, their computational complexity in terms of the amount of required coefficients and their retrieval computational cost is presented.

The rest of the paper is organised as follows. In Section 2, the four shape descriptors are described. In Section 3, their indexing methodology and efficiency is presented. In Section 4, the shape-based descriptors are compared and evaluated against each other in both terms of retrieval and computational complexity performance. The paper is concluded in Section 5.

2 Shape descriptors

In this section, we describe four important shape descriptors: FD, curvature scale space descriptors, angular radial transform

(ART) descriptors and image moment descriptors. FD and curvature scale space descriptors are contour-based since they are extracted from the contour, while image moments and ART descriptors are region-based extracted from the whole shape region.

2.1 Fourier descriptors

FD have been successfully applied to many shape representation applications, especially to character recognition. Their nice characteristics, such as simple derivation, simple normalisation and its robustness to noise, have made them very popular in a wide range of applications [9]. The FD are obtained through Fourier transform on a complex vector derived from shape boundary coordinates (x_n, y_n), $n = 0, 1, \ldots, N-1$. The complex vector $\bar{U}$ is given by the difference of the boundary points from the centroid (x_c, y_c) of the shape

$$\bar{U} = \begin{pmatrix} x_0 - x_c + \mathrm{i}(y_0 - y_c) \\ x_1 - x_c + \mathrm{i}(y_1 - y_c) \\ \vdots \\ x_n - x_c + \mathrm{i}(y_n - y_c) \end{pmatrix}, \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where

$$x_c = \frac{1}{N}\sum_{n=0}^{N-1} x(n), \quad y_c = \frac{1}{N}\sum_{n=0}^{N-1} y(n) \tag{2}$$

The centroid subtraction, which represents the position of the shape from boundary coordinates, makes the representation invariant to translation. One-dimensional Fourier transform is then applied on $\bar{U}$ to obtain the Fourier transformed coefficients

$$\bar{F}_k = FFT[\bar{U}] \tag{3}$$

The magnitudes of the coefficients $|\bar{F}_k|$ are normalised by the magnitude of the first coefficient $|\bar{F}_0|$ for scaling invariance. The acquired FD are translation, rotation and scale invariant.

By limiting the number of coefficients k, we are able to reduce the high-frequency noise to a great extent, leaving at the same time the main details of the patterns. Thus, the application of limited number of FD has the effect of lowpass filtering. However, this reduction leads also to loss of spatial information in terms of fine detail.

2.2 Curvature scale space descriptor

The curvature scale space descriptor (CSS) treats shape boundary as a 1D signal, and analyses this signal in scale space [10]. Since curvature is a very important local measure on how fast a planar contour is turning, the curvature scale space is exploited. Thus, the zero crossings of curvatures at different scales, through Gaussian smoothing, are useful for shape description because they represent the perceptual features of shape contour.

The CSS descriptors are obtained after scale normalisation on a complex vector derived from shape boundary coordinates (x_n, y_n), $n = 0, 1, \ldots, N-1$. Then the curvature is derived from shape boundary points as follows [11]

$$k(t) = (\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t))/(\dot{x}^2(t) + \dot{y}(t)^2)^{3/2} \tag{4}$$

where $\dot{x}(t)$, $\dot{y}(t)$ and $\ddot{x}(t)$, $\ddot{y}(t)$ are the first and the second derivatives at location t, respectively. Curvature zero-cross points are then located in the shape boundary. The shape is then evolved into the next scale by applying Gaussian smoothing

$$x'(t) = x(t) * g(t, \sigma, c), \quad y'(t) = y(t) * g(t, \sigma, c) \tag{5}$$

where $*$ means convolution, and $g(t, \sigma, c)$ is the Gaussian function as follows

$$g(t, \sigma, c) = \mathrm{e}^{[-(t-c)^2]/2\sigma^2} \tag{6}$$

With c consant and as σ increases, the evolving shape becomes smoother. New curvature zero-crossing points are located at each scale. This process continues until no curvature zero-crossing points are found. The acquired zero-crossing points are then plotted onto the (t, σ) plane to create the CSS contour map.

The peaks, or the local maxima of the CSS contour map (only peaks higher than a threshold are considered) are then extracted out and sorted in descending order of σ. After normalisation, the CSS peaks are used as CSS descriptors to index the shape.

The descriptor is translation invariant. Scale invariance is achieved by normalising all the shapes into a fixed number of boundary points. An efficient number of points is 128 as proposed in [6]. Since rotation of shape causes circular shifting of CSS peaks on the t-axis, the rotation invariance is achieved by circular shifting the highest peak (primary peak) to the origin of the CSS map. The final descriptor consists of a vector of 128 coefficients representing the boundary points, and its values correspond to the normalised peaks.

2.3 Angular radial transform

The ART is a moment-based image description method adopted in MPEG-7 as a 2D region-based shape descriptor [12]. The ART is a complex orthogonal unitary transform defined on a unit disc based on complex orthogonal sinusoidal basis functions in polar coordinates. The ART coefficients, F_{nm} of order n and m, are defined by

$$F_{mn} = \int\int V^*_{m,n}(x, y) f(x, y) \mathrm{d}x\, \mathrm{d}y \tag{7}$$

where $f(x,y)$ is an image function in polar co-ordinates and $V_{m,n}(x, y)$ is the ART basis function that is separable along the angular and radial directions

$$V_{m,n}(r, \theta) = R_n(r) A_m(\theta) \tag{8}$$

with

$$A_m(\theta) = \frac{1}{2\pi}\mathrm{e}^{\mathrm{j}m\theta} \tag{9}$$

and

$$R_n(r) = \begin{cases} 1, & (n = 0) \\ 2\cos(\pi n r), & (n > 0) \end{cases} \tag{10}$$

The ART descriptor is defined as a set of normalised magnitudes of the set of ART coefficients. Rotational invariance is obtained by using the magnitude of the coefficients. In order to achieve translation invariance, the center of the polar coordinate

doi: 10.1049/iet-ipr.2009.0246

system is defined as the center of mass of the object, which can be easily acquired by geometric moments [13]

$$\bar{x} = \frac{\int\int f(x, y)x \, \mathrm{d}x \, \mathrm{d}y}{\int\int f(x, y)\mathrm{d}x \, \mathrm{d}y} \tag{11}$$

$$\bar{y} = \frac{\int\int f(x, y)y \, \mathrm{d}x \, \mathrm{d}y}{\int\int f(x, y)\mathrm{d}x \, \mathrm{d}y} \tag{12}$$

According to the MPEG-7 standard, the ART descriptor is expressed by 140 bits, consisting of 35 four-bit coefficients, since it is defined that the calculated coefficients have orders ($n < 3$, $m < 12$), normalised by $|F_{00}|$. Therefore the normalised scaling invariant coefficients are given by

$$\bar{F}_{mn} = \frac{2\pi F_{mn}}{\int_s f(x, y)\mathrm{d}x \, \mathrm{d}y} \tag{13}$$

2.4 Image moments

Image moments (IM) have been proved applicable in various recognition tasks. The chosen image moment are not invariant only under translation, rotation and scaling of the object but also under general affine transformation [14]. The affine moment invariants are derived by means of the theory of algebraic invariants and more specifically by means of decomposition of affine transformation into six one-parameter transformations. The six affine invariants used are defined below

$$\begin{aligned}
I_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{11}^2 - \mu_{11}^2) \\
I_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - \mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 \\
&\quad + 4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
I_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
&\quad + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \\
I_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{21}\mu_{02}\mu_{03} \\
&\quad + 9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{03}\mu_{21} \\
&\quad + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\
&\quad - 8\mu_{11}^3\mu_{03}\mu_{30} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} + 9\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}^2\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2) \\
I_5 &= \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) \\
I_6 &= \frac{1}{\mu_{00}^9}(\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{40}\mu_{13}^2 \\
&\quad - \mu_{04}\mu_{31}^2 - \mu_{22}^3)
\end{aligned} \tag{14}$$

where μ_{pq} is defined by

$$\mu_{pq} = \int\int_{\text{object}} f(x, y)(x - x_t)^p(y - y_t)^q \mathrm{d}x \, \mathrm{d}y \tag{15}$$

The proposers of the moment-based method have suggested the use of either all six or only the first four invariant moments for the description of objects. We used the first case for all our experiments.

3 Indexing

The indexing process for each descriptor was performed based on their optimal number of coefficients as described in the previous section based on the current literature. More specifically for FD we used the first 32 descriptors, for curvature scale space descriptors we used 128 feature points, for the ART the first 35 descriptors and the six-image moment descriptors as follows

$$\mathbf{f}_{\text{FD}} = \left\{\frac{|\bar{F}_1|}{|\bar{F}_0|}, \frac{|\bar{F}_1|}{|F_0|}, \ldots, \frac{|\bar{F}_{32}|}{|\bar{F}_0|}\right\} \tag{16}$$

$$\mathbf{f}_{\text{CSS}} = \{C_1, C_2, \ldots, C_{128}\} \tag{17}$$

$$\mathbf{f}_{\text{ART}} = \{\bar{F}_{11}, \bar{F}_{21}, \ldots, \bar{F}_{112}\} \tag{18}$$

$$\mathbf{f}_{\text{IM}} = \{I_1, I_2, \ldots, I_6\} \tag{19}$$

However, the quantisation of the values of the descriptor coefficients depends on the descriptor itself. The quantisation is a critical issue in retrieval systems since the amount of the processed data depends mainly on the size of the encoded database. For FD an eight-bit representation is adequate for each coefficient. For curvature scale space descriptors, the required bits for peak representation after optimisation and relational quantisation are 9 [15]. The four-bit representation for ART coefficient is accepted by the MPEG-7 standard [12]. Finally, for image moments, since are up to the third order, a double precision floating point representation of 64-bit is required [16].

4 Comparison results

We compared the performance of each descriptor on two datasets. The first dataset is the MPEG-7 data set for the Core Experiment CE-Shape-1, part B, illustrated in [17]. We used the 1400 image dataset divided in 70 shape classes of 20 images each. Sample dataset images are illustrated in Fig. 1*a*. This set is very useful for testing similarity-based retrieval and the shape descriptors' robustness to various arbitrary shape distortions.

The second data set consists of 200 affine transformed bream fish and 1100 marine fish, which are unclassified consisting the MPEG-7 data set for the Core Experiment CE-Shape-1, part C. The 200 bream fish are designated as queries. Sample dataset images are illustrated in Fig. 1*b*. This set is for testing the shape descriptors robustness to non-rigid object distortions.

The size of both encoded databases based on the proposed indexing are shown in Fig. 2. It can be seen that ART and FD involve less amount of data for offline encoding of both databases. CSS is the most expensive among all the four shape descriptors. The same comparison results are derived when comparing the descriptors based on their computational

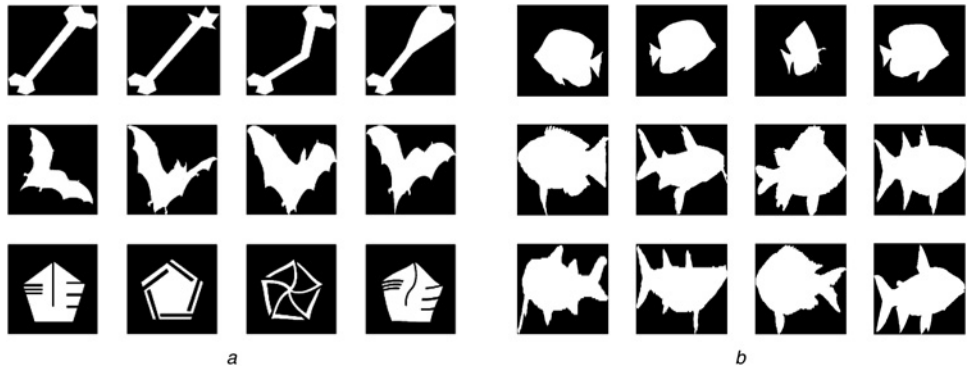

Fig. 1 *Sample images of the MPEG-7 dataset CE-Shape-1*

a Part B
b Part C

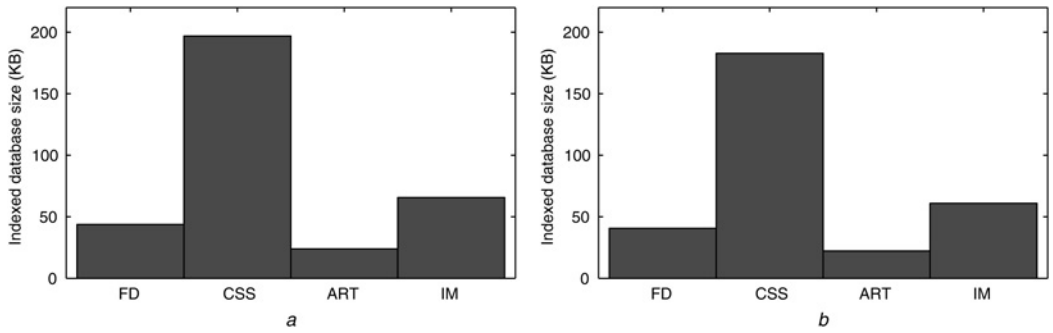

Fig. 2 *Required data for indexing Core Experiment CE-Shape-1 image database*

a Part B
b Part C

efficiency in retrieval when using an Athlon-1.2 GHz with 1 GB memory. The average time taken for retrieval on Part-B shape database is given in Table 1. For online matching, CSS is the most expensive, while FD and IM are much more efficient in terms of the average retrieval time. ART presents the lowest average time of retrieval for each query.

For performance measure we used the precision and recall of the retrieval for the evaluation of the query results. Precision P is defined as the ratio of the number of retrieved relevant shapes r to the total number of retrieved shapes n, $P = r/n$. Precision P measures the accuracy of the retrieval and the speed of the recall. Recall R is defined as the ratio of the number of retrieved relevant images r to the total number m of relevant shapes in the whole database, $R = r/m$. Recall R measures the robustness of the retrieval performance.

In order to study the behaviour of each shape descriptor, we first present their retrieval performance in two different shapes. The first shape is shown in Fig. 3*a* representing a glass with great boundary variation. The second shape is a circular device with fine details along its circular shape as shown in Fig. 4*a*. From the precision-recall charts it can be seen that FD is more robust to general boundary variations while CSS fails to retrieve objects with no prominent contours. IM presents a fine overall performance; however, when compared with the results of both shapes, it can be seen that its superiority is met in the circular device since it can capture the fine details of the contour. The ART presents an average performance in both shapes, realising an acceptable trade-off between accuracy and complexity.

Table 1 Time for retrieval using the FD, CSS, ART and IM

Shape descriptor	Total time of retrieval, ms	Average time of retrieval for each query, ms
FD	34 025	24.3
CSS	129 367	92.4
ART	19 526	13.9
IM	61 287	43.7

The average precision and recall result are presented in Fig. 5, and Fig. 6, for Part B and Part C shape databases, respectively. From the average results, IMs are presenting the best retrieval performance in all the datasets tested. It is clear from Fig. 5, that for simple shape transformation such as scaling, rotation, and affine transform, IMs have much higher performance than the other descriptors. They also demonstrate robust performance in generic shape variations as shown in Fig. 6. On average, their retrieval results are more perceptually acceptable than FDs. However, moments because of their nature, have an intrinsic problem in shapes with relatively large stretching. FDs which are contour-based are also robust to such irregularities of boundary as shown in retrieval results using the Part B database. ART in both datasets has not the best description accuracy; however, is a small and simple descriptor with fast retrieval answer, applicable where only a low order of computational complexity is required. CSS robustness in boundary variations is very limited, thus it is proposed to be combined with global descriptors in order to form a more robust shape descriptor.

doi: 10.1049/iet-ipr.2009.0246

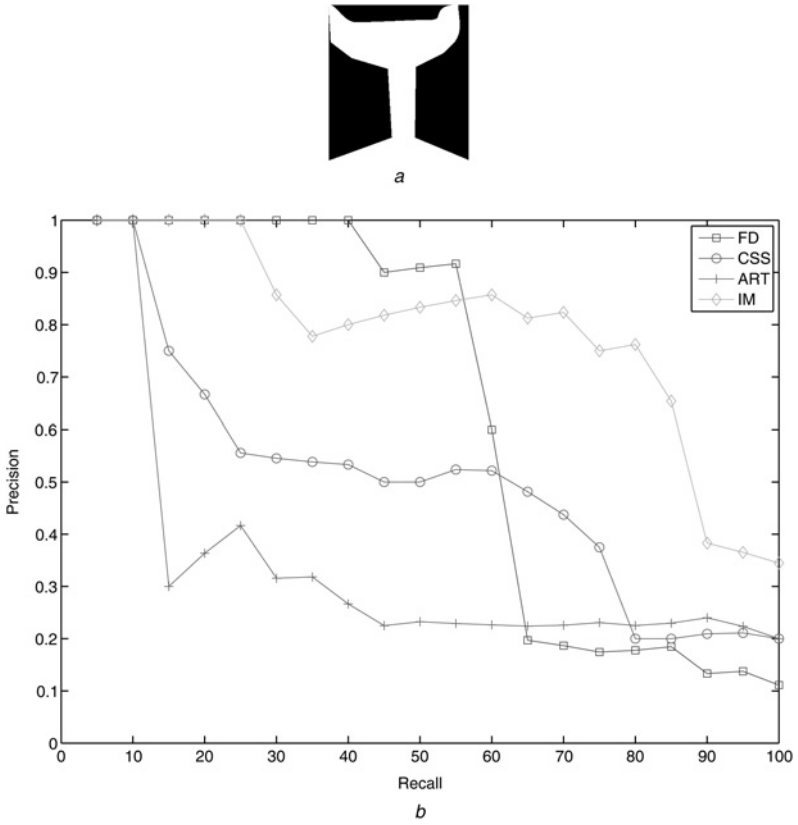

Fig. 3 *Retrieval results using FD, CSS, ART and IM for glass-18*

a Query image
b Precision/recall graph for the query image

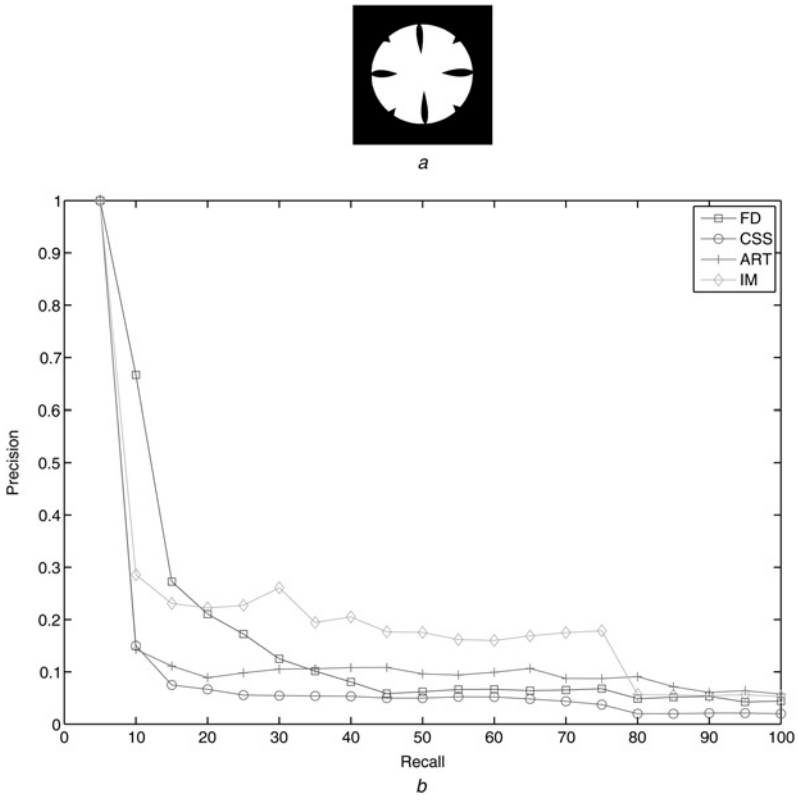

Fig. 4 *Retrieval results using FD, CSS, ART and IM for device9–15*

a Query image
b Precision/recall graph for the query image

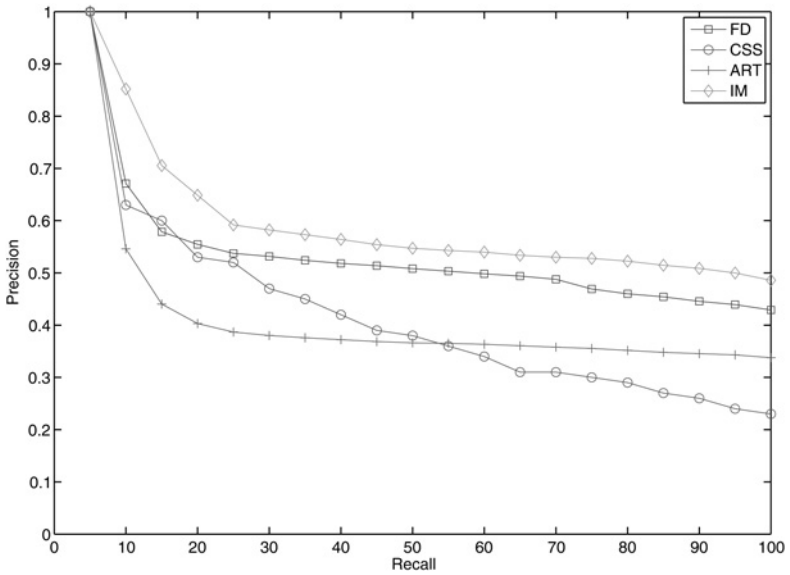

Fig. 5 *Average precision and recall of retrieval using FD, CSS, ART and IM on MPEG-7 contour shape database Part B*

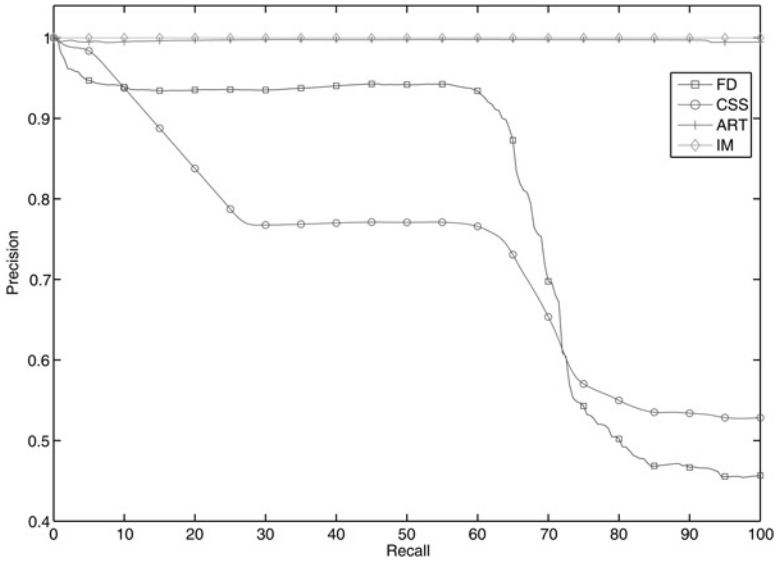

Fig. 6 *Average precision and recall of retrieval using FD, CSS, ART and IM on MPEG-7 contour shape database part C*

5 Conclusion

In this paper, we have made a study and a comparison on four shape descriptors for shape retrieval. Results show that in terms of scaling, translation, and rotation invariance and compactness, image moment descriptors obtain more credits than the other three methods. Although image moment descriptors lacks the contour information reflected in contour based descriptors, retrieval results favour its performance. ART is a small and simple descriptor with fast retrieval answer presenting the lowest average time of retrieval, making it suitable where only a low order of computational complexity is required.

6 Acknowledgment

This work was supported, in part, by a project Archimedes-III contract.

7 References

1 Persoon, E., Fu, K.: 'Shape discrimination using Fourier descriptors', *IEEE Trans. Syst. Man Cybern.*, 1997, **7**, (3), pp. 170–179
2 Belongie, S., Malik, J., Puzicha, J.: 'Shape matching and object recognition using shape contexts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (4), pp. 509–522
3 Berretti, S., Del Bimbo, A., Pala, P.: 'Retrieval by shape similarity with perceptual distance and effective indexing', *IEEE Trans. Multimedia*, 2000, **2**, (4), pp. 225–239
4 Tangelder, J., Veltkamp, R.: 'A survey of content based 3D shape retrieval methods', *Multimedia Tools Appl.*, 2008, **39**, (3), pp. 441–471
5 Paquet, E., Rioux, M., Murching, A., Naveen, T., Tabatabai, A.: 'Description of shape information for 2-D and 3-D objects', *Signal Process., Image Commun.*, 2000, **16**, (1–2), pp. 103–122
6 Zhang, D., Lu, G.: 'Evaluation of MPEG-7 shape descriptors against other shape descriptors', *Multimedia Syst.*, 2003, **9**, (1), pp. 15–30
7 Latecki, L., Lakamper, R., Eckhardt, U.: 'Shape descriptors for non-rigid shapes with a single closed contour'. IEEE Conf. on Computer Vision and Pattern Recognition, 2000, vol. 1
8 Zhang, D., Lu, G.: 'Review of shape representation and description techniques', *Pattern Recognit.*, 2004, **37**, (1), pp. 1–19
9 Zhang, D., Lu, G.: 'Shape-based image retrieval using generic Fourier descriptor', *Signal Process., Image Commun.*, 2002, **17**, (10), pp. 825–848

doi: 10.1049/iet-ipr.2009.0246

10 Jalba, A., Wilkinson, M., Roerdink, J.: 'Shape representation and recognition through morphological curvature scale spaces', *IEEE Trans. Image Process.*, 2006, **15**, (2), pp. 331–341
11 Zhang, D., Lu, G.: 'A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval', *J. Vis. Commun. Image Represent.*, 2003, **14**, (1), pp. 39–57
12 Bober, M.: 'MPEG-7 visual shape descriptors', *IEEE Trans. Circuits Syst. Video Technol.*, 2001, **11**, (6), pp. 716–719
13 Kotoulas, L., Andreadis, I.: 'An efficient technique for the computation of ART', *IEEE Trans. Circuits Syst. Video Technol.*, 2008, **18**, (5), pp. 682–686
14 Flusser, J., Suk, T.: 'Pattern recognition by affine moment invariants', *Pattern Recognit.*, 1993, **26**, (1), pp. 167–174
15 Mokhtarian, F., Bober, M.: 'Curvature scale space representation: theory, applications, and MPEG-7 standardization' (Kluwer Academic Publishers, 2003)
16 Kotoulas, L., Andreadis, I.: 'Efficient hardware architectures for computation of image moments', *Real-Time Imag.*, 2004, **10**, (6), pp. 371–378
17 Latecki, L., Lakamper, R., Eckhardt, T.: 'Shape descriptors for non-rigid shapes with a single closed contour'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2000, vol. 1