IST 2009 - International Workshop on Imaging Systems and Techniques
Shenzhen, China
May 11-12, 2009

A Comparative Study of Invariant Descriptors for Shape Retrieval

A. Amanatiadis[1], V.G. Kaburlasos[2], A. Gasteratos[3], and S.E. Papadakis[2]

[1]Laboratory of Electronics, Department of Electrical and Computer Engineering,
Democritus University of Thrace, 12 Vas. Sofias Str., GR-67100, Xanthi, Greece
Email: aamanat@ee.duth.gr

[2]Division of Computing Systems, Department of Industrial Informatics,
Technological Educational Institution of Kavala, GR-65404, Kavala, Greece
Email: {vgkabs, spap} @teikav.edu.gr

[3]Laboratory of Robotics and Automation, Department of Production and Management Engineering,
Democritus University of Thrace, University Campus Kimmeria, GR-67100, Xanthi, Greece
E–mail: agaster@pme.duth.gr

***Abstract*—This paper presents a comparative study between scale, rotation and translation invariant descriptors for shape representation and retrieval. Specifically, we studied Fourier, angular radial transform and image moment descriptors for shape representation. Since shape is one of the most widely used image feature exploited in content-based image retrieval systems, we studied for each descriptor, the number of coefficients needed for indexing and their retrieval performance. Results showed that moment descriptors present the best performance in both terms of shape representation quality as well as in the amount of required coefficients.**

***Index Terms*—Shape representation, image retrieval system, shape matching, invariant descriptors.**

I. Introduction

Shape representation compared to other features, like texture and color, is much more effective in semantically characterizing 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]. Pattern recognition by shape descriptors can be used in many practical tasks, for example in image matching, multitemporal image sequence analysis, shape classification and character recognition.

Various shape descriptors exist in the literature, mainly categorized 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 necessarily rely on shape boundary information, but they do not reflect local features of a shape. Therefore, for generic purposes, both types of shape representations are necessary.

II. Shape Descriptors

In this section we describe three important shape descriptors: Fourier descriptors, angular radial transform descriptors and image moment descriptors. Fourier descriptors are contour-based while image moments and angular radial transform descriptors are region-based.

A. Fourier Descriptors

Fourier descriptors have been successfully applied to many shape representation applications, especially to character recognition. Their nice characteristics, such as simple derivation, simple normalization, and its robustness to noise, have made them very popular in a wide range of applications [3]. The Fourier descriptors 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

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

where

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

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

$$\bar{F_k} = FFT[\bar{U}] \quad (3)$$

The magnitudes of the coefficients $|\bar{F}_k|$ are normalized by the magnitude of the first coefficient $|\bar{F}_0|$ for scaling invariance. The acquired Fourier descriptors 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 Fourier descriptors has the effect of lowpass filtering. However, this reduction leads also to loss of spatial information in terms of fine detail.

B. Angular Radial Transform

The angular radial transform (ART) is a moment-based image description method adopted in MPEG-7 as a 2D region-based shape descriptor [4]. The ART is a complex orthogonal unitary transform defined on a unit disk based on complex orthogonal sinusoidal basis functions in polar coordinates. The ART coefficients, F_{nm} of order n and m, are defined by

$$F_{mn} = \int\int V^*_{m,n}(x,y) f(x,y) dxdy \tag{4}$$

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

$$V_{m,n}(r,\theta) = R_n(r) A_m(\theta) \tag{5}$$

with

$$A_m(\theta) = \frac{1}{2\pi} e^{jm\theta} \tag{6}$$

and

$$R_n(r) = \begin{cases} 1, & (n = 0) \\ 2\cos(\pi n r), & (n > 0) \end{cases} \tag{7}$$

The ART descriptor is defined as a set of normalized 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 system is defined as the center of mass of the object, which can be easily acquired by geometric moments [5]

$$\bar{x} = \frac{\int\int f(x,y) x dxdy}{\int\int f(x,y) dxdy} \tag{8}$$

$$\bar{y} = \frac{\int\int f(x,y) y dxdy}{\int\int f(x,y) dxdy} \tag{9}$$

According to the MPEG-7 standard, the ART descriptor is expressed by 140 bits, consisting of 35 4-bit coefficients, since it is defined that the calculated coefficients have orders ($n < 3, m < 12$), normalized by $|F_{00}|$. Therefore, the normalized scaling invariant coefficients are given by

$$\bar{F}_{mn} = \frac{2\pi F_{mn}}{\int_s f(x,y) dxdy} \tag{10}$$

C. Image Moments

Image moments have been proved applicable in various recognition tasks. The chosen image moment are not invariant only only under translation, rotation and scaling of the object but also under general affine transformation [6]. 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:

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

where μ_{pq} is defined by

$$\mu_{pq} = \int\int_{object} f(x,y)(x - x_t)^p (y - y_t)^q dxdy \tag{12}$$

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.

III. Indexing

The indexing process for each descriptor was performed based on their optimal number of coefficients described in literature. More specifically for Fourier descriptors we used the first 32 descriptors, for the angular radial transform the first 35 descriptors and the 7 image moment descriptors as follows:

$$\mathbf{f}_{fourier} = \{\frac{|\bar{F}_1|}{|\bar{F}_0|}, \frac{|\bar{F}_1|}{|\bar{F}_0|}, \ldots, \frac{|\bar{F}_{32}|}{|\bar{F}_0|}\} \tag{13}$$

$$\mathbf{f}_{art} = \{\bar{F}_{1\,1}, \bar{F}_{2\,1}, \ldots, \bar{F}_{11\,2}\} \tag{14}$$

$$\mathbf{f}_{moment} = \{I_1, I_2, \ldots, I_6\} \tag{15}$$

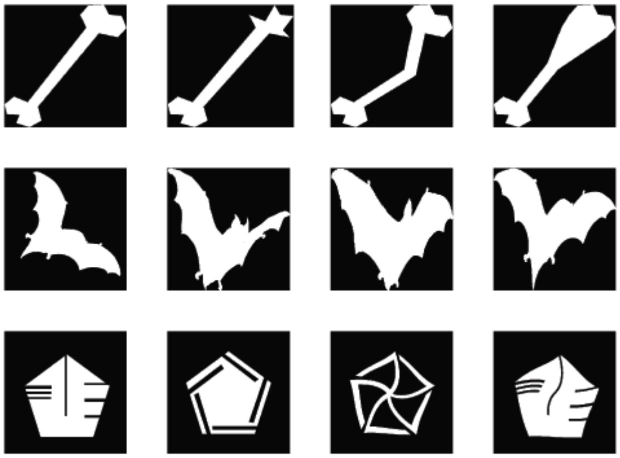


Fig. 1. Sample images of the MPEG-7 data set.

IV. Comparison Results

We compared the performance of each descriptor on the MPEG-7 data set for the Core Experiment CE-Shape-1, part B, illustrated in [7]. We used the $1,400$ image data set divided in 70 shape classes of 20 images each. One image from each shape class was used as a query image and the 10 of total 20 first relevant images were retrieved. Sample dataset images are illustrated in Fig. 1.

The Euclidean distance between two descriptor vectors was used for similarity measurement as follows:

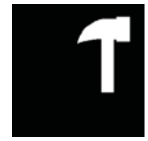
$$dist = (\sum_{i}^{N} |f_q^i - f_d^i|^2)^{1/2} \tag{16}$$

where N is the number of used descriptors for indexing the shapes, $\mathbf{f}_q$ the query descriptor vectors, and $\mathbf{f}_d$ the database descriptor vectors.

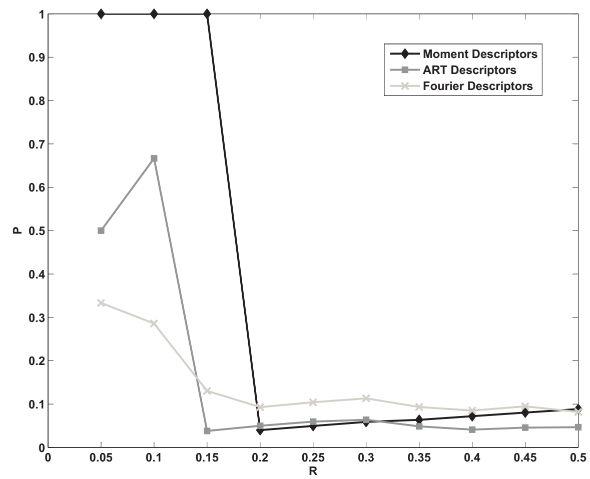
For performance measure we used the precision and recall of the retrieval for the evaluation of the query results. Precision P is defined as the ratio of the number of retrieved relevant shapes r to the total number of retrieved shapes n, $P = r/n$. Precision P measures the accuracy of the retrieval 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. The precision and recall results for 4 query images are depicted in Figure 2 to Figure 5, with moment descriptors presenting the best average retrieval performance. However, it is known that the retrieval performance of shape descriptors depends highly on the shape database. Thus, the comparison results between the different descriptors can provide a better measure of efficiency.

V. Conclusion

In this paper we have made a study and a comparison on three shape descriptors for shape retrieval. Results show that in terms of scaling, translation, and rotation invariance and compactness, image moment descriptors get more credits than



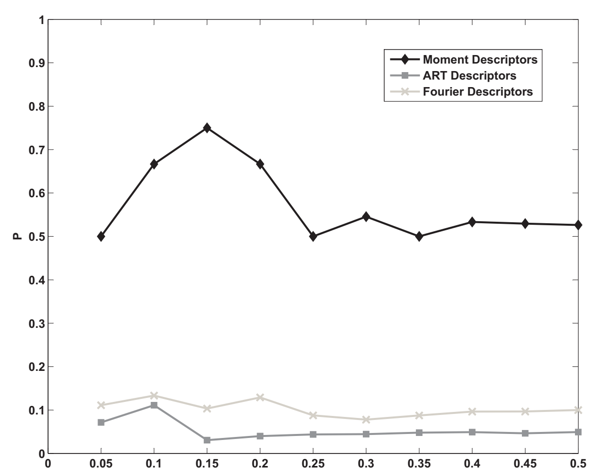
(a)



(b)

Fig. 2. Retrieval results for *hammer-15*: (a) query image; (b) Precision/Recall graph for the query image.

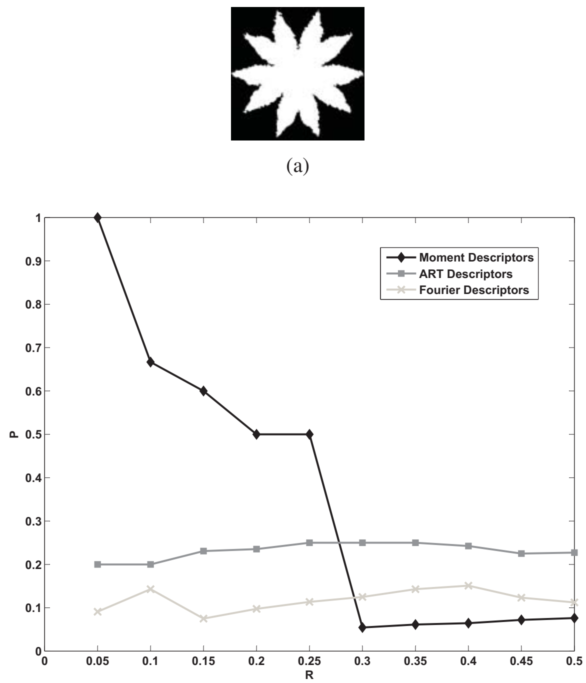
(a)



(b)

Fig. 3. Retrieval results for *bottle-14*: (a) query image; (b) Precision/Recall graph for the query image.

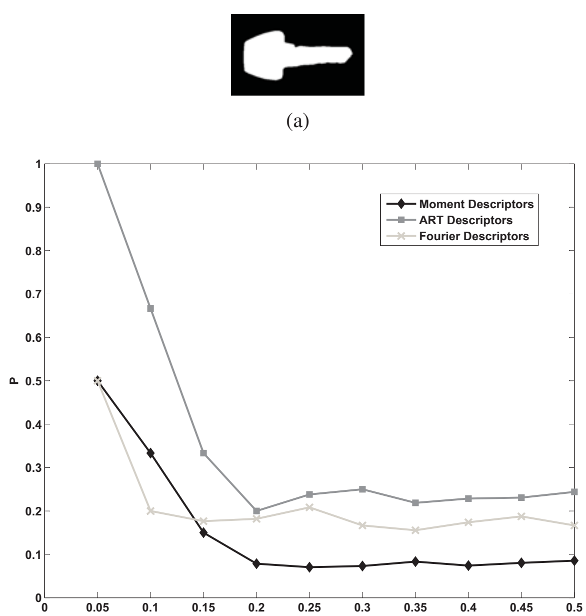
the other two methods. Although image moment descriptors lacks the contour information reflected in Fourier descriptors, retrieval results favor its performance. Furthermore, while angular radial transform captures strong perceptual shape features, Fourier descriptors are more robust to general boundary variations than angular radial transform.



(a)

(b)

Fig. 4. Retrieval results for *device7-13*: (a) query image; (b) Precision/Recall graph for the query image.



(a)

(b)

Fig. 5. Retrieval results for *key-19*: (a) query image; (b) Precision/Recall graph for the query image.

REFERENCES

[1] E. Persoon and K. Fu, "Shape discrimination using Fourier descriptors," *IEEE Trans. Systems, Man and Cybernetics*, vol. 7, no. 3, pp. 170–179, 1977.

[2] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 509–522, 2002.

[3] D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.

[4] M. Bober, "MPEG-7 visual shape descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716–719, 2001.

[5] L. Kotoulas and I. Andreadis, "An efficient technique for the computation of ART," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 5, pp. 682–686, 2008.

[6] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, 1993.

[7] L. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 2000.