

Image Registration Based On Statistical Descriptors In Frequency Domain

Min-hyuk Chang¹, Muhammad Bilal Ahmad², Cheul-hee Lee³, Jong-hoon Chun⁴,
Seung-jin Park⁵ and Jong-an Park¹

¹ Electronics Information & Communications Eng., Chosun Univ., #375 Seosuk-dong, Kwangju, 501-759, Korea
Tel.: +82-62-230-7064, Fax.: +82-62-232-6776

² Signal & Image Processing Lab., Dept. of Mechatronics Kwangju Institute of Science and Technology (KJIST), Korea

³ Dept. of Electronics and Information & Communication Eng., Kochang Polytechnic Collage, Korea

⁴ Dept. of Information & Communication Eng., Provincial Collage of Namdo, Korea

⁵ Dept. of Biomedical Eng., Chounnam Univ., Hospital, Kwangju, Korea
e-mail : millcre@chosun.ac.kr, japark@chosun.ac.kr, bilal@geguri.kjist.ac.kr,

lch@kopo.or.kr, owl7026@kornet.net, sjinpark@cnuh.com

Abstract: Shape description and its corresponding matching algorithm is one of the main concerns in MPEG-7. In this paper, a new method is proposed for shape registration of 2D objects for MPEG-7. Shapes are recognized using the Hu statistical moments in frequency domain. The Hu moments are moment-based descriptors of planar shapes, which are invariant under general translation, rotational, scaling, and reflection transformation. The image is transformed into frequency domain using Fourier Transform. Annular and radial wedge distributions for the power spectra are extracted. Different statistical features (Hu moments) are found for the power spectrum of each selected transformed individual feature. The Euclidean distance of the extracted moment descriptors of the features are found with respect to the shapes in the database. The minimum Euclidean distance is the candidate for the matched shape. The simulation results are performed on the test shapes of MPEG-7.

1. Introduction

An incommensurable amount of audiovisual information is becoming available in digital form, in digital archives, on the World Wide Web, in broadcast data streams and in personal and professional databases, and this amount is only growing. The value of information often depends on how easy it can be found, retrieved, accessed and filtered and managed. In spite of the fact that users have increasing access to these resources, identifying and managing them efficiently is becoming more difficult, because of the sheer volume.

MPEG-7, formally named "Multimedia Content Description Interface", aims to create a standard for describing the multimedia content data that will support some degree of interpretation of the information's meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes shall support as broad a range of applications as possible.

In other words: MPEG-7 will specify a standard set of Descriptors that can be used to describe various types of multimedia information. To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features (or 'descriptors') will be extremely useful. MPEG-7 will make

the web as searchable for multimedia content as it is searchable for text today.

MPEG-7 visual description tools consist of basic structures and descriptors that cover the basic visual features such as color, texture, shape, motion, localization, etc. Each category consists of elementary and sophisticated descriptors.

In this paper, the topic of shape description is presented, and a new method is proposed for shape registration of 2D objects. There are four types of shape descriptors: object region-based shape, contour-based shape, 3D shape, and 2D-3D-Multiple view. The proposed method will only use region based shape descriptors for shape registration.

Shape analysis is useful in a number of applications of machine vision, including medical image analysis, aerial image analysis, and manufacturing. Shape description and its corresponding matching algorithm is one of the main concerns in MPEG-7. Shape retrieval consists of two stages: Shape description, and Shape matching. In this paper we will recognize objects by shape matching using the statistical invariant descriptors in frequency domain.

After the objects with closed contours have been detected, a corresponding mechanism must be established to match these objects irrespective of the change of size, position, or orientation. One of the useful shape descriptors is based on the theory of moments. The moment invariants are moment-based descriptors of planar shapes, which are invariant under general translation, rotational, scaling, and reflection transformation. The lower-order invariants, which are composed of the second- and third-order central moments, are usually sufficient for most registration tasks on remotely sensed image matching.

The notion of object shape, while intuitively clear, may have many meanings. Firstly, most of the real-world objects are 3-D and a 3-D-shape descriptor has been developed by MPEG-7 [1, 2, 3]. However, the image and video world usually deals with 2-D projections (onto an image plane) of real-world objects, and MPEG-7 also provides tools to describe such "2-D" shapes.

Even in the 2-D case, there can be two notions of similarity. This is shown in Figure 1. Objects in the first row have similar spatial distribution of pixels and are therefore similar according to a region-based criterion. However, they clearly have different outline contours. When contour-

based similarity is concerned, objects shown in each column are similar. Posing a query with the object located in the first row and second column will result in retrieved object from the first row (if region-based similarity is concerned) or second column (if contour-based similarity is concerned). MPEG-7 supports both notions of similarity using region-based and contour-based shape descriptors.

Statistical invariant descriptors in spatial domain will give similar results for objects in the same row. However, the statistical invariant descriptors give different values in the frequency domain.

In this paper, frequency transform-based statistical method is used. The normalized frequency transform is invariant to translation, rotation and scaling. The image is transformed into frequency domain using Fourier Transform. The DC component is set at the middle of the transformed image using Fourier Shift property. Power Spectrum is the square of the magnitude. Annular and radial wedge distributions for the power spectra are extracted. Bilinear interpolation is used at the points where the values are undetermined. The ring boundaries can be set at arbitrary frequencies, such as (0-1), (1-2), (2-4), (4-8), (8-16), (16-32), (32-64). The annular and radial wedges can be set arbitrarily. Different statistical features (Hu moments [5]) are found for the power spectrum of each selected transformed individual feature. The Euclidean distance of the extracted moment descriptors of the features are found with respect to the shapes in the database. The minimum Euclidean distance is the candidate for the matched shape. The simulation results are performed on the test shapes of MPEG-7. The recognition rate of the proposed method is higher than the spatial-based methods.

This paper is organized as follows. Section 2 describes the different statistical descriptors that will be utilized for shape similarity measures. Section 3 explains the proposed algorithm of statistical descriptors in frequency domain. The simulation results are shown in section 4. At the end, we will conclude our paper with few final remarks.

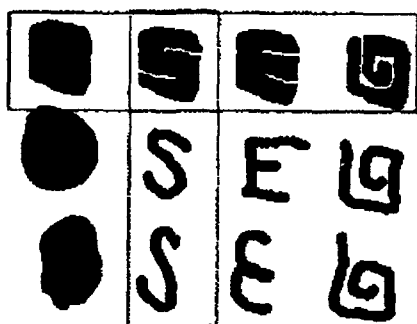


Figure 1. Examples of contour- and region-based shape similarity.

2. Statistical Descriptors

After the objects with closed contours have been detected, a corresponding mechanism must be established to match these objects irrespective of the change of size, position, or orientation. One of the useful shape descriptors is based on the theory of moments [2,4,5]. The moment invariants are

moment-based descriptors of planar shapes, which are invariant under general translation, rotational, scaling, and reflection transformation. Central moments and scaling-invariant moment representation can be employed to produce a set of invariant moments which are further invariant to rotational and reflection differences. The lower-order invariants, which are composed of the second- and third-order central moments, are usually sufficient for most registration tasks on remotely sensed image matching. We can use statistical invariants to recognize objects, because they provide simple means of comparison and provide position and orientation information (i.e. Moments etc.). They are too simple for some applications, and do not give a unique means of identification. Statistical moments work directly with regions of pixels in the image. Such measures are usually represented by a single value. The statistical descriptors find: area, length, perimeter, elongation and Moments of Inertia, etc. The moments of a binary image $b(x, y)$ are given by

$$\mu_{pq} = \sum_x \sum_y b(x, y) x^p y^q, \quad (1)$$

where p and q define the order of moment. Where $b(x, y)$ can be omitted as it has only 1 and 0 values, so sums are only taken where $b(x, y)$ has values 1, and the equation reduces to

$$\mu_{pq} = \sum_x \sum_y x^p y^q, \quad (2)$$

The center of gravity of the object can be found from moments as:

$$\bar{x} = \frac{\mu_{10}}{\mu_{00}}, \quad \bar{y} = \frac{\mu_{01}}{\mu_{00}}, \quad (3)$$

where (\bar{x}, \bar{y}) are the coordinates of the center of gravity. The pq th discrete central moment m_{pq} of a region is defined by

$$m_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q \quad (4)$$

where the sums are taken over all points (x, y) . Zero-order moment m_{00} represents the binary object area. Second order moments express the distribution of matter around the center of gravity, and are called moments of inertia. We can form seven new moments from the central moments that are invariant to changes of position, scale and orientation of the object represented by the region using central moments of lower orders. For moments of order up to three, these are:

$$\begin{aligned} M1 &= m_{20} + m_{02} \\ M2 &= (m_{20} - m_{02})^2 + 4m_{11}^2 \\ M3 &= (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2 \\ M4 &= (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2 \\ M5 &= (m_{30} - 3m_{12})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] \\ &\quad + (3m_{21} - m_{03})(m_{21} + m_{03})[3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2] \end{aligned}$$

$$\begin{aligned}
M6 &= (m_{20} + m_{02})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] + 4m_{11}(m_{30} \\
&\quad + m_{12})(m_{03} + m_{21}) \\
M7 &= (3m_{21} - m_{03})(m_{12} + m_{30})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] \\
&\quad - (m_{30} - 3m_{12})(m_{12} + m_{03})[3(m_{30} + m_{12})^2 - (m_{21} + \\
&\quad m_{03})^2]
\end{aligned} \tag{5}$$

All moments $M1, M2, M3, \dots, M7$ are translation, rotational and scale invariants. These invariants will help us in the object recognition. The principal axes of inertia that define a natural coordinate system for a region. Let θ be the angle that the x-axis of the natural coordinate system (the principal axes) makes with the x-axis of the reference coordinate system. Then θ is given by

$$\theta = \frac{1}{2} \tan^{-1} \left[\frac{2m_{11}}{m_{20} - m_{02}} \right] \tag{6}$$

From the principal axes of inertia, we can find the orientation of the objects.

3. Power Spectrum

For shape matching, we used transform-based method [6]. This method was based on the power spectrum of the Fourier transformation. The Fourier transformation has two components, namely, magnitude and phase. The power spectrum is the square of the magnitude. The Fourier transform is a complex function

$$F(u, v) = R(u, v) + I(u, v) \tag{7}$$

The power spectrum is simply the square of the absolute magnitude of the transform

$$P(u, v) = |F(u, v)|^2 = R^2(u, v) + I^2(u, v) \tag{8}$$

For discrete digital images the integration is replaced by summation. Since digital images, in general, and regions of digital images, in particular, are of finite extent, the summation need only be computed over those integer coordinates where the image or region is defined. We used bilinear approximation values for the non-integer regions.

After taking the Fourier transform of the image, the Fourier transform is partitioned into different areas of interest. The partition is done in angle and radius. Figure 2. shows the partitioning of the Fourier domain for the case of five different spatial frequencies and eight directions.

The DC component of the Fourier transform is shifted to the center in the frequency domain using the shift technique for centering the DC component of the Fourier transform. The different annular and radial wedge distributions are made arbitrarily at different directions and angles. The ring boundaries are at the frequencies:

(1) [0-1]; (2) [1-2]; (3) [2-4]; (4) [4-8]; (5) [8-16];

The wedges are at every 45 degrees. The ring boundaries and wedge angles are arbitrarily changed according to the desired number of features.

For shape matching, different features are selected in frequency domain, as shown in Figure 2., marked with small lines. Power spectrum is found for those selected

regions. For the non-integer position, bilinear approximations are used. The Hu statistical descriptors (Equation 5.) as discussed in section 2, are found for the power spectrum of the selected region.

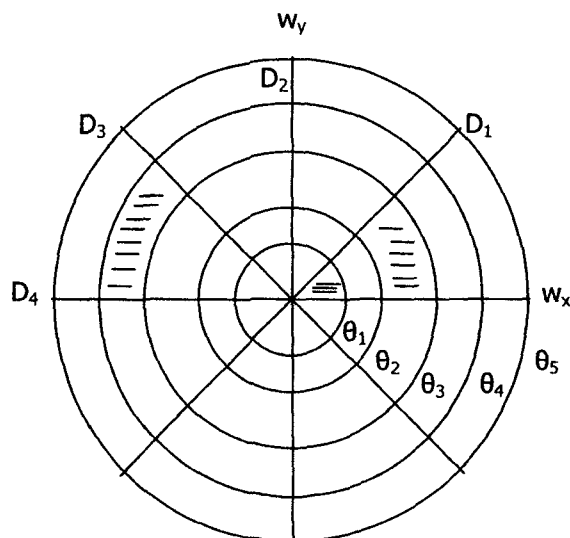


Figure 2. Fourier Domain Partitioning (Annular and radial wedge distribution).

For matching criterion, we used Minkowsky Distance and Euclidean Distance criterions.

A. Minkowsky Distance Matching Criterion

This measure is defined as the sum of the absolute differences between two image feature vectors, one corresponding to the query's shape, Q , and the other to a shape in the contents' database, C , for which a description is available.

$$D_M = \sum_{k=0}^{N-1} |Q[k] - C[k]|, \tag{9}$$

where N is the number of image features. Once computed the distances for all images, the final distance will be the minimum value distance for matching criterion.

$$D = \min_i \{D_{Mi}\} \tag{10}$$

where i corresponds to the i th image in the database.

B. Euclidean Distance Matching Criterion

The Euclidean distance is one of the most immediate, simplest and most frequently used similarity measures between two vectors. This measure is defined as the square root of the sum of the squared differences between two feature vectors, one belonging to the query's shape, Q , and the other belonging to a shape in the contents' database, C , for which a description is available.

$$D_E = \sqrt{\sum_{k=0}^{N-1} (Q[k] - C[k])^2}, \tag{11}$$

The Hu statistical descriptors are found for the selected power spectrum of the regions in the partitioned Fourier domain (Figure 2.). The Minkowsky distances, or

Euclidean distances are found for the query image with respect to database images. The minimum distance gives the most closet shape matched.

4. Simulation Results

For simulation, we performed different experiments on MPEG-7 test images. Few MPEG-7 test images are shown in Figure 3.

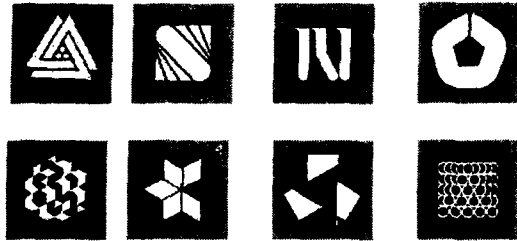


Figure 3. MPEG-7 Binary Test images

The experiments are performed in both spatial and frequency domain. The simulation results shows better results in the frequency domain for the Hu statistical descriptors. Table 1. shows the different simulation results on the MPEG-7 test images. The images are divided into three categories. The first category is called "Simple Objects", which are closed objects with no holes. The second category is called "Complex Objects", which are closed objects but contains small holes. The performance results are carried on Hu Statistical descriptors in the spatial domain and in the frequency domain. Table 1. used Minkowsky distances matching criterion for image matching, while Table 2. used Euclidean distance matching criterion.

Table 1. Shape Matching performance using Minkowsky Distance Criterion on MPEG-7 test images

Method's Name	Simple Objects	Complex Objects	Average Performance
Spatial Domain	85.0%	70 0%	78.5%
Proposed Method (Frequency Domain)	90.0%	82.0 %	86.0%

Table 2. Shape Matching performance using Euclidean Distance Criterion on MPEG-7 test images

Method's Name	Simple Objects	Complex Objects	Average Performance
Spatial Domain	92.0%	75 0%	84 5%
Proposed Method (Frequency Domain)	95.0%	85.0 %	90.0 %

From table 1. and 2., it is shown that the shape matching performance is enhanced by using Hu Statistical descriptors for the power spectrum of the selected regions in the partitioned Fourier domain. Further, the Euclidean distance matching criterion gives better results as compared to the Minkowsky distance matching criterion.

5. Conclusions

In this paper, we have proposed shape matching using Hu statistical descriptors in the frequency domain. The Hu descriptors give lower performance in the spatial domain as compared to the frequency domain. The images are transformed in frequency domain using the Fourier transform. The Fourier transform is partitioned in both annular and radial directions. The power spectrum is found for the different arbitrarily selected features in the Fourier partitioned. The Hu descriptors are found for the power spectrum of the selected features. Different statistical features (Hu moments) are found for the power spectrum of each selected transformed individual feature, and then matched with the database objects.. The simulation results are performed on the test shapes of MPEG-7. The results are better in the frequency domain as compared to the spatial domain.

References

- [1] A. Goshtasby, "Registration of images with geometric distortions," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, pp. 60-64, Jan. 1988.
- [2] Morton Nadler, Eric P. Smith, *Pattern Recognition Engineering*, John Wiley & Sons Inc., 1993.
- [3] L. G. Brown, "A survey of image registration techniques," *Computer Survey*, vol. 24, no. 4, pp. 325-376, 1992.
- [4] R.C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1993Allen.
- [5] M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inform. Theory*, vol. IT-8, no. 2, pp. 179-187, 1962.
- [6] G. Lendaris and G. Stanley, "Diffraction Pattern Sampling for Automatic Pattern Recognition," *Proceedings of the IEEE*, vol. 58, pp. 198-216, 1970.