MAXIMAL DISK BASED HISTOGRAM FOR SHAPE RETRIEVAL*

*Wai-Pak Choi , Kin-Man Lam** and Wan-Chi Siu*

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

In this paper, we propose a robust and efficient representation scheme for shape retrieval, which is based on the normalized maximal disks used to represent the shape of an object. The maximal disks are extracted by means of a fast skeletonization technique with a pruning algorithm. The logarithm of the radii of the normalized maximal disks is used to construct a histogram to represent the shape. The retrieval performance of this maximal disk based histogram approach is compared to other methods, including moment invariants, Zernike moments, and curvature scale-space. Experimental results show that our proposed representation scheme outperforms the other methods under affine transformation and different noise levels.

Index Terms--Shape descriptor, skeletonization, maximal disk, histogram distance, shape retrieval.

1. INTRODUCTION

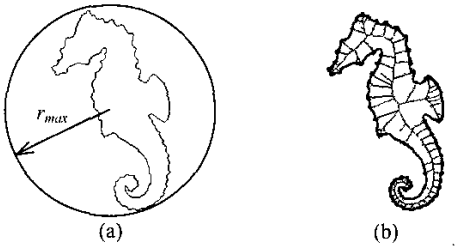
In designing a content-based image retrieval system [1,2], retrieval efficiency and accuracy are the two most important issues to be considered, especially for large multimedia databases. Research on content-based image retrieval has attracted great interest and attention over the past decade. Feature selection and extraction is the most important step for image retrieval. To describe different types of images, different visual features, such as shape, texture, color, etc., should be considered.

Shape retrieval is one of the most challenging areas covered by the literature that deals with retrieval based on shape similarity. There are different traditional approaches for representing the shape characteristics, such as Fourier descriptors [3,4], moment invariants [5-7], Zernike Moments [8,9], etc. However, these methods are sensitive to noise and distortion. The curvature scale-space (CSS) method [10-12] is one of the most successful contour-based features used as a shape descriptor, and has been a shape descriptor in the MPEG-7 standard [13]. The representation of the curvature scale-space has been shown to be robust under the similarity transformations such as scaling, orientation changes, translation, and even shearing and noise. However, the method is easily affected by distortion or occlusion. A shape descriptor which is compact, robust to noise and distortion, and provides an accurate representation of an object's shape is highly desirable.

In this paper, a robust and efficient representation scheme based on the distribution of the maximal disks used to represent a shape is proposed. The maximal disks of a shape are extracted by a fast skeletonization technique [14] with a pruning algorithm [15]. The robust feature is formed by the logarithm of the radii of the normalized maximal disks of an object. A histogram based on the distribution of the radii is constructed and used for shape matching. In Section 2, we present our proposed shape descriptor in details. Experimental results are illustrated in Section 3, where the performance of our proposed scheme is compared to other shape representation schemes based on moment invariants, Zernike moments, and curvature scale-space.

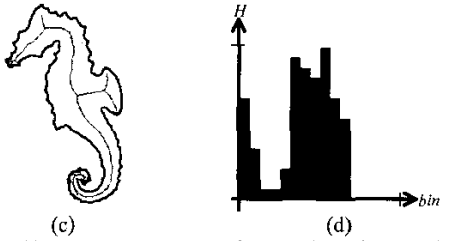
2. A NEW SHAPE DESCRIPTOR

The shape of an object can be obtained by contour extraction or image segmentation methods, such as the adaptive snake method [16] or any edge follower method. The extracted contour or shape is then encircled completely by a minimum disk, as shown in Fig. 1(a). The radius of this minimum disk is denoted as r_{max}. In order to obtain a descriptor independent of scale, the shape is normalized such that the radius r_{max} is equal to a fixed value $N/2$. In other words, the object forms an image of size $N \times N$. Based on the extracted contour, the skeleton of the shape is then extracted using a fast skeletonization method with a pruning algorithm. The skeletonization method uses the loci of those maximal disks inside an object to represent its skeleton. In our scheme, the distribution of these radii are used to represent the object. Figure 1(b) shows the result after skeletonization, which contains many branches. This is due to the discrete representation of the shape. In order to remove the branches, a pruning algorithm is applied, and the pruned result is shown in Fig. 1(c). The radii of those maximal disks associated with the pruned skeleton are then used to form a histogram for shape representation. The corresponding histogram for the skeleton in Fig. 1(c) is shown in Fig. 1(d).



(a) (b)

*This work was supported by research grant G-V596 from The Hong Kong Polytechnic University, Hong Kong.
**E-mail:enkmlam@polyu.edu.hk



(c) (d)

Figure 1: Shape representation of a marine shape: (a) the contour, (b) the skeleton, (c) the pruned skeleton, and (d) the maximal disk based histogram.

2.1. Maximal disk based histogram

Similar shapes will result in similar histograms for the radii of the maximal disks. The maximal disks are extracted by our skeletonization algorithm. Each representative radius of the maximal disks and its corresponding percentage form a pair of attributes that can be used to describe the shape characteristics in an image. Basically, the histogram is formed by the maximal disks along the skeleton of a shape, so it is independent of the translation and orientation of the object. The radii of those maximal disks used to represent the object are normalized by the radius r_{max}, so histogram formed on the basis of these normalized radii is also invariant to the object's scale. Maximal disks with a larger radius capture the global shape, while those with a smaller radius represent the fine details. As the number of maximal disks with small radii is much more than that of large radii, the allocation of the normalized radii to the histogram bins should be performed in a non-linearly manner. Therefore, the radius r of a maximal disk is normalized by r_{max} and then quantized to a bin value, $bin(r)$, of the histogram as follows:

$$bin(r)=\frac{1}{2}\log_{10}\left(\frac{r}{r_{max}}\times 100\right)\times N_{bin} \qquad (1)$$

where N_{bin} is the maximum number of bins and $bin(r)$ is in the range of $[1, \ldots, N_{bin}]$. The histogram of a shape with maximal disks of radii in the range of $[1, R]$, where $R \le r_{max}$, is defined as follows:

$$H(bin(r))=\frac{count(bin(r))}{n} \quad \text{where} \quad r=1,2,\ldots,R \qquad (2)$$

where $count(bin(r))$ represents the number of maximal disks assigned to $bin(r)$ of the histogram, n is the total number of maximal disks for the object, and $\sum H(bin(r)) = 1$.

In order to have a compact representation of the shape descriptor, the histogram is set to have 16 bins, and each bin is represented by 4 bits. Therefore, 64 bits are used to represent a histogram or a shape. The 4-bit value is used to represent the number of maximal disks with a particular range of radii in the corresponding bin. The maximal disk based histogram of a marine shape is shown in Fig. 1(d).

2.2. Histogram Comparison

The difference between two histograms is computed by using the quadratic-form distance measure [17,18], which takes the cross-bin similarity into account when measuring the distance. Suppose that the maximal disk based histograms of two images, P and Q, are represented as $\mathbf{H}_p = (H_p(1),\ldots H_p(N_{bin}))^T$, and $\mathbf{H}_q = (H_q(1),\ldots H_q(N_{bin}))^T$, respectively. The quadratic-form distance measure D is then defined as follows:

$$D(P,Q) = \sqrt{(\mathbf{H}_p-\mathbf{H}_q)^T \mathbf{A}(\mathbf{H}_p-\mathbf{H}_q)} \qquad (3)$$

where $\mathbf{A}=[a_{ij}]$ is a similarity matrix that incorporates the cross-bin similarity, and a_{ij} denotes the similarity between bins i and j.

Each bin of the histogram corresponds to a particular range of logarithm radii of the normalized maximal disks, and the similarity matrix represents the correlation between the bin under consideration and its neighbors. The similarity matrix $\mathbf{A}$ used is defined as follows:

$$a_{ij} = \exp\left(-\left(\frac{i-j}{\sigma}\right)^2\right) \qquad (4)$$

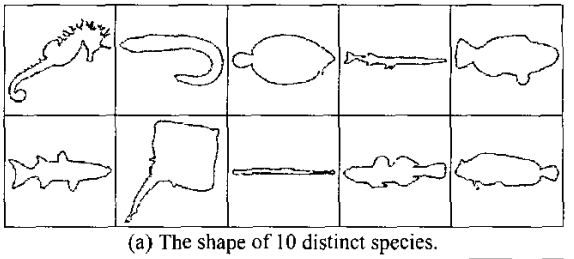
where $i, j = [1,\ldots,N_{bin}]$ and σ is the variance of the similarity parameter. The value of this distance measure D can then be used to determine the similarity between the two shapes.

3. EXPERIMENTAL RESULTS

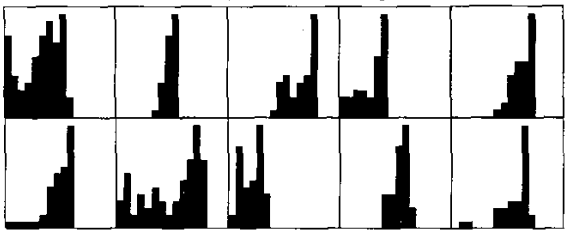
The retrieval performance of our proposed algorithm is compared to the traditional algorithms including Zernike moments, moment invariants and curvature scale-space. Part of the SQUID database (Shape Queries Using Image Databases) [11] was used in the experiments. The effects of affine transformation, noise, and distortion on the performances of the different algorithms were investigated and evaluated. The experiments were conducted on a Pentium 4 2.4GHz PC.

3.1. Generation of image databases

In our experiments, 10 distinct species of marine creatures were selected from the SQUID database. Each of the marine images is represented as a set of closed-contour points, then converted to binarized images, and normalized to a size of 256×256. The 10 different species form 10 classes, and each class contains at most 8 similar species. In our system, this makes up a database consisting of 76 images in all, which is similar to the databases constructed in [11]. One shape from each of the classes and its corresponding maximal disk based histogram are shown in Fig. 4. We can observe that the histograms of distinct species are different.



(a) The shape of 10 distinct species.



(b) The corresponding maximal disk based histograms.

Figure 4: The 10 representative shapes used in the experiments.

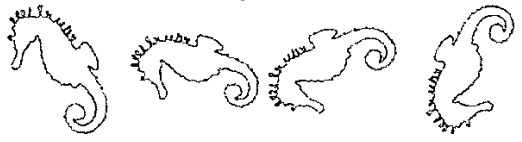
In order to evaluate the accuracy of our proposed algorithm, three types of databases were generated by applying different affine transformations, different levels of noise added to the contour points, and different distortions added to the contour points in the original database. The first database was generated by using scaling factors between 0.8 and 0.9, and rotated by an arbitrary angle. The second database was generated by using 10 levels of noise between 0.0 and 18.0. The third database was generated by randomly adding different distortions to the contour points of the marine images. Hence, the number of classes in each of these databases is the same. The query set used in the experiments was formed by all the images within the database. The runtime required for extracting the proposed shape descriptors for 760 images in each of the databases is approximately 24 seconds, and the average runtime for retrieval for each query is about 168 ms.

3.2. Comparisons of different shape descriptors

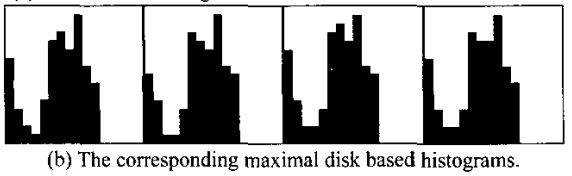
In order to visualize the robustness of the different shape descriptors, the effects of affine transformation, noise, and distortion are investigated in the following.

3.2.1. Effect of affine transformations

Figure 5 shows the maximal disk based histograms with the object under different orientations and scales. It can be observed that the histograms of the images are very similar to each other.



(a) The simulated images with different orientations and scales.



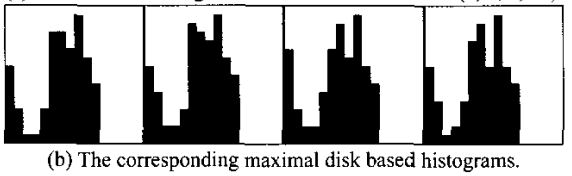
(b) The corresponding maximal disk based histograms.

Figure 5: The maximal disk based histograms of a marine shape with different orientations and scales.

3.2.2. Effect of noise levels

Figure 6 shows the histograms using our proposed representation scheme for a shape under different levels of noise. It can be observed that the histograms of the images with different noise variances are similar to each other.

(a) The simulated images with different noise levels (4, 6, 8, 10).

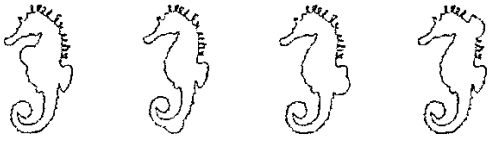


(b) The corresponding maximal disk based histograms.

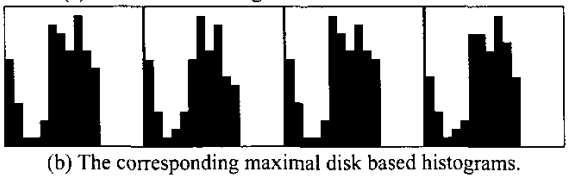
Figure 6: The maximal disk based histograms of a marine shape under different noise levels.

3.2.3. Effect of distortions

Figure 7 shows the maximal disk based histograms under different distortions. It can be observed that the histograms of the images with different distortions are changed but the overall representations of the histograms are still similar to each other.



(a) The simulated images with different distortions.



(b) The corresponding maximal disk based histograms.

Figure 7: The maximal disk based histograms of a marine shape under different distortions.

3.3. The retrieval performances

To measure the retrieval performance of the different shape representation schemes, the precision and recall rates for the different retrieval approaches are measured. The precision rate and recall rate are defined as follows:

$$\text{Precision Rate} = \frac{\text{number of relevant images selected}}{\text{total number of retrieved images}} \quad (5)$$

$$\text{Recall Rate} = \frac{\text{number of relevant images selected}}{\text{total number of similar images in the database}} \quad (6)$$

In our experiments, the best-matched 80 marine images were retrieved from the different databases for each query image. The average values of the precision rate and recall rate were computed and plotted as precision-recall graphs. The Zernike moments, the moment invariants, and the curvature scale-space were compared to our proposed feature by means of the precision-recall graphs.

Figure 8 shows the precision-recall graphs of the different algorithms based on the three databases. The performance of the Zernike moments falls dramatically when the shapes are rotated and scaled, as compared with that of the other approaches. The performance of the moment invariants under noise is dropped when compared with others, due to its sensitivity to noises. The performance of the curvature scale-space when the shapes are distorted also falls dramatically compared to the others. Our proposed scheme has the best retrieval performances under affine transformation and different noise levels. Figure 8(c) shows that our proposed algorithm still outperforms the moment invariants and the curvature scale-space, but is only very slightly interior to that of the Zernike moments.

4. CONCLUSIONS

This paper proposed a robust and efficient representation scheme for representing and retrieval shapes. Our algorithm is based on the histogram of the logarithmic radii of the normalized maximal disks of a shape. The centers of these maximal disks are located along the skeleton of the shape. This representation scheme is compact and can achieve a good retrieval performance level. The performance of our proposed scheme is superior to the moment

invariants, Zernike moments, and curvature scale-space methods when the shapes are under affine transformation, and at different noise levels. Nevertheless, its performance is very slightly interior to that of the Zernike moments when the shapes are distorted.

5. REFERENCES

[1] Y. S. Kim, W. Y. Kim, "Content-based trademark retrieval system using a visually salient feature", *Image and Vision Computing*, vol. 16, pp. 931-939, 1998.

[2] B. Gunsel and A. M. Tekalp, "Shape similarity matching for query-by-example", *Pattern Recognition*, vol. 31, no. 7, pp. 931-944, 1998.

[3] A. Blumenkrans, "Two-Dimensional Object Recognition using a Two-Dimensional Polar Transform", *Pattern Recognition*, vol. 24, no. 9, pp. 879-890, 1991.

[4] B.G. Quinn, "Estimating frequency by interpolation using Fourier coefficients", *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1264-1268, May 1994.

[5] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", *IRE Trans. Inform. Theory*, vol. IT-8, pp. 179-187, Feb. 1962.

[6] T. H. Reiss, "The revised fundamental theorem of moment invariants", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, Aug 1991.

[7] C. C. Chen, "Improved Moment Invariants for shape discrimination", *Pattern Recognition*, vol. 26, no. 5, pp. 683-686, 1993.

[8] A. Wallin and O. Kubler, "Complete sets of complex Zernike moment invariants and the role of the pseudoinvariants", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1106-1110, Nov. 1995.

[9] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489-497, May 1990.

[10] M. Bober, "Shape descriptor based on curvature scale space", Lancaster, U.K., *MPEG-7 proposal* P320, Feb. 1998.

[11] F. Makharian and S. Abbasi, "Shape similarity retrieval under affine transforms", *Pattern Recognition*, vol. 35, pp. 31-41, 2002.

[12] F. Mokhtarian and A. K. Mackworth, "A theory of multiscale, curvature-based shape representation for planar curves", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 789-805, Aug. 1992.

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

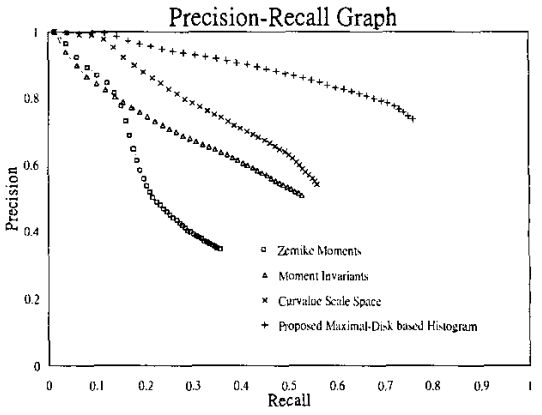
[14] W. P. Choi, K. M. Lam and W. C. Siu, "Extraction of an Euclidean Skeleton based on a connectivity Criterion", accepted by *Pattern Recognition*, 2002.

[15] W. P. Choi, K. M. Lam and W. C. Siu, "An Efficient and Accurate Algorithm for extracting a Skeleton", *Proceedings. 15th International Conference on Pattern Recognition*, vol. 3, pp. 750-753, 2000.

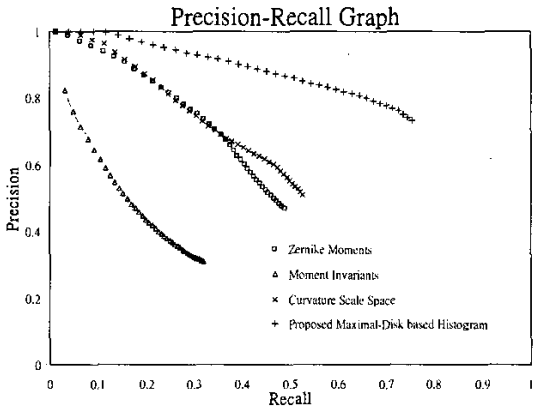
[16] W. P. Choi, K. M. Lam and W. C. Siu, "An adaptive active contour model for highly irregular boundaries", *Pattern Recognition*, vol. 34, no. 2, pp. 323-331, 2001.

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

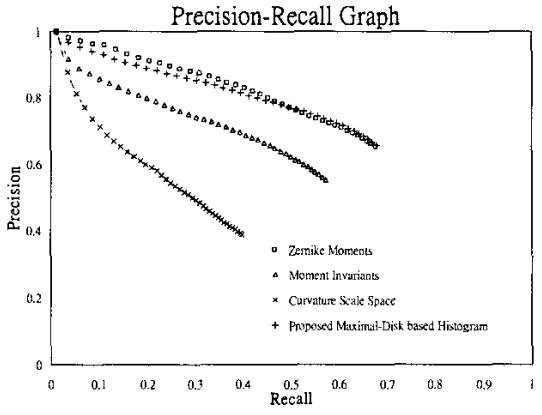
[18] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729 -736, July 1995.



(a) Shapes are rotated and scaled.



(b) Noises are added to the shapes.



(c) Shapes are distorted.

Figure 8: Comparison of the precision-recall graphs with 80 retrieval images using the three different shape databases: (a) affine transformed, (b) different noise levels, and (c) different distortions.