AN EFFICIENT SHAPE REPRESENTATION AND DESCRIPTION TECHNIQUE

Y. Ebrahim, M. Ahmed, S.C. Chau,

Wilfrid Laurier University
Physics and Computer Science
Waterloo, ON, Canada

W. Abdelsalam

University of Guelph
Computer Information Sciences
Guelph, ON, Canada

ABSTRACT

In this paper we present a novel approach to shape representation and description based on the combination of the Hilbert space filling curve and Wavelet analysis. Our objective is to capitalize on the localization-preserving nature of the Hilbert space filling curve and the approximation power of the Wavelet transform. The object image is scanned using the Hilbert curve and the resulting vector is smoothed using the wavelet transform and sampled. The technique is O(N) for both representation and comparison. We present some experimental results on the MPEG-7 dataset, Kimia-99 dataset, ETH-80 dataset, and a logo dataset.

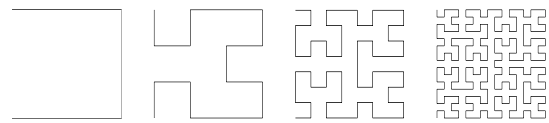
***Index Terms*—** Image shape analysis, Pattern recognition, Wavelet transforms.

1. INTRODUCTION

As the preferred visual feature for image retrieval, shape has attracted a great deal of attention in the image representation and retrieval research community. Although many shape representation and description techniques have been reported [**?**, 1–5], there is still a need for ones that possess the MPEG-7 criteria of good retrieval accuracy, compact features, general application, low computation complexity, robust retrieval performance, and hierarchical representation [6]. This paper introduces a new region-based shape representation and description technique. The idea is to capture the shape features by the Hilbert Curve (HC) producing a 1D version of the image which is smoothed and sampled to produce the image's Shape Feature Vector (SFV). Not only is the proposed technique invariant to translation, scaling, and stretching, but also robust to occlusion and articulation. Experimental results are given on a number of benchmark datasets to show the efficacy of the proposed technique.

2. THE HILBERT SPACE FILLING CURVE

A space filling curve is a continuous path, which visits every point in a 2-dimensional grid exactly once and never crosses itself. Space-filling curves provide a linear order of the points of a grid. The goal is to do so while keeping the points that are close in 2-dimensional space close together in the linear order. The Hilbert curve is one of the most popular space filling curves in use because of its excellent localization capabilities. Figure 1 shows a Hilbert curve at 4 different levels of detail.



(a) (b) (c) (d)

Fig. 1. Four levels of the Hilbert filling curve.

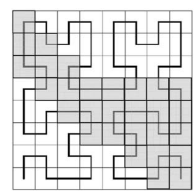


Fig. 2. HC superimposed on shape.

3. USING THE HC FOR SHAPE REPRESENTATION AND DESCRIPTION

To capture the variation between foreground and background pixels, which has the shape information embedded in it, the image is scanned by using the HC, and the intensity value for each visited pixel is saved in a vector, V. The first half of V for the shape in Figure 2 is 000000000111000001111101000001. The locality preserving nature of the HC results in a vector that reflects the clustering of pixels in the image. To smooth out noise and keep the main shape features intact, the wavelet transform is applied to V, producing the vector WV. Then, WV is then sampled to obtain the vector SWV which is normalized to produce the object's Shape Features Vector (SFV). Figure 3 depicts these steps. At search time, the distance be-

1-4244-1437-7/07/$20.00 ©2007 IEEE

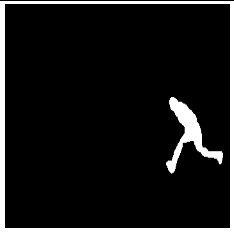
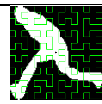
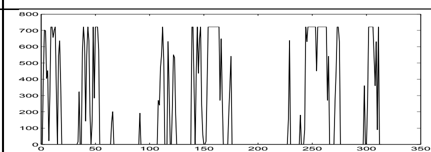
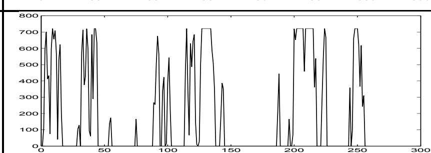
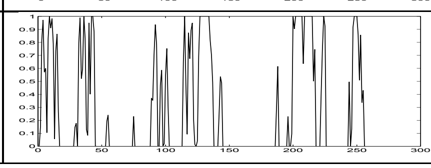
Step	Sample output
Original image	
Crop and scan	
V	
Approximate	
Sample	
Normalize	

Fig. 3. Steps of the proposed shape representation and description approach.

tween the search and the database images' SFVs is computed by a distance measure.

4. ALGORITHM PARAMETERS

There are a number of parameters associated with the proposed algorithm. These variables are the Hilbert curve level, the wavelet used, the wavelet approximation level, and the SFV size.

4.1. HC Level

For a HC to visit each pixel in an image the image must be of size 2^nx2^n, where n is the level of the curve. Instead of having to normalize the image to a 2^nx2^n size, the HC level is set at the largest n such that $2^n \leq min(I, J)$, where I and J are the dimensions of the image. To minimize the effect of missing the unvisited pixels (i.e., those which do not coincide

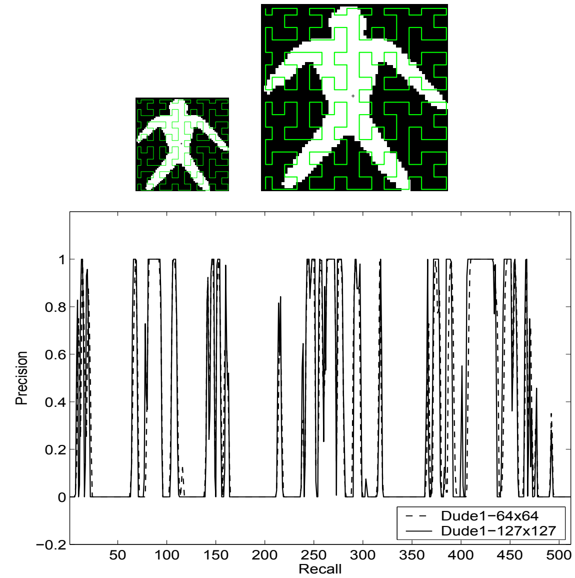


Fig. 4. SFVs for 64x64 and 127x127 versions of the image with a cross-correlation = 0.94

with HC vertices), the HC is *stretched* such that the unvisited pixels are distributed evenly on both dimensions.Figure 4(a) reflects the SFVs of an image normalized to a size of 2^nx2^n (left) and $2^n + (2^n - 1)$ x $2^n + (2^n - 1)$ (right), where $n = 6$. Notice that both images are scanned with a level 6 HC. The HC superimposed on the images is of a lower level for clarity purposes. Although the HC had to be stretched considerably to cover the right image, the two SFVs are almost identical with a very high cross-correlation of 0.94. This shows that the technique is virtually invariant to scaling.

4.2. Choosing a Wavelet

Before a wavelet transform is applied to a dataset, a wavelet must be chosen from a variety of wavelets. For this purpose, wavelets from the Daubechies, Biorthogonal, Reverse Biorthogonal, Coiflets, and symlets families are examined. To choose the most appropriate wavelet, based on empirical evidence, the Kimia-99 dataset is represented by different wavelets from the aforementioned wavelet families. The resultant representations are queried for each image in the dataset. Also, the Precision Recall (PR) curve for each dataset is created and the results are compared.

The following wavelets are identified as the most appropriate: db1, db2, db3, coif1, sym2, sym3, sym4, sym5, sym6, bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, rbio1.1, rbio1.3, rbio1.5, rbio2.2, and rbio2.4. The average retrieval precision achieved by these wavelets is comparable. Due to the simplicity of the db1 wavelet, it is selected for all the experiments in this paper.

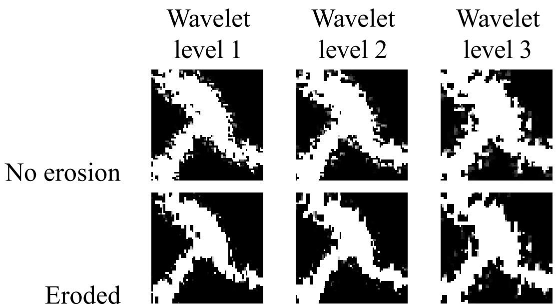


Fig. 5. Original image reconstruction from a 1024 point SFV with and without erosion.

4.3. Choosing the Wavelet Approximation Level

In this work, the wavelet transform is selected as a means of smoothing the SFV such that only the most pertinent features of the shape are retained. Too much approximation, however, can result in a loss of valuable information. Empirical evidence shows that a wavelet approximation level of 3 produces the best results.

4.4. Determining the SFV Size

Because the SFV size affects the sampling rate, the larger the SFV is, the more details are preserved and the more the expense to store and compare SFVs. This typical trade-off between efficiency and information preservation is not linear, however. The effect of the information, gained by increasing the SFV size, is often much less significant than the increase in the SFV size.

Because the SFV is a washed down 1D representation of the image, an approximation of the original image is obtained by mapping the SFV back to 2D. The larger the SFV, the closer the approximation is to the original image. This feature is crucial in distributed database environments, where the SFVs and the actual images are stored on different servers. Typically, the SFVs are located where the search is done, and the search results are retrieved from the central database. The ability to see an approximation of the search results allows the user to identify the images of interest to be retrieved from the central database. This, of course, reduces the load both on the network and the DBMS.

Figure 5 depicts the reconstructions obtained from a 256 and a 1024 representations at wavelet levels 1, 2, and 3. As expected, the bigger the SFV and the smaller the wavelet level, the closer the reconstructed image is to the original. To show how the resulting image can be processed to reduce noise, the same figure shows the same set of images after they are eroded.

Table 1. Retrieval rate for the MPEG-7 dataset using three SFV sizes.

SFV size	64	128	256	512	1024
	63.6%	72.4%	74.5%	75.5%	75.7%

5. EXPERIMENTAL RESULTS

A number of experiments are described in this section. The proposed technique is tested on the MPEG-7, Kimia-99, ETH-80, and Logo datasets. Note that the MPEG-7 and Kimia-99 classes with rotations are preprocessed by rotating each image such that its major axis is horizontal.

If not indicated otherwise, the following parameter values are used: SFV size 256, wavelet approximation level 3, db1 wavelet, and cross-correlation similarity measure.

5.1. MPEG-7 Dataset

Table 1 lists the results of the MPEG-7 core experiment (CE-Shape-1) part B for the proposed technique with SFV sizes 64, 128, 256, 512, and 1024. As expected, the retrieval rate improves as the SFV size grows. As the SFV size grows, less information is lost at the sampling stage which allows for a more accurate representation of the object. However, this comes at a cost in terms of both storage space and execution time.

5.2. Kimia-99 Dataset

Figure 6 signifies the effect of the wavelet approximation level on the retrieval accuracy, represented by the PR curves. The approximation level is varied between 0 (no approximation) and 5. From Figure 6, it is clear that the wavelet approximation level affects the retrieval accuracy. The best results seem to be associated with the approximation levels 3 and 4. This suggests that the smoothing effect of the wavelet approximation at these levels produces the best *quality* SFVs. This results in maintaining the shape features, and smoothing away noise and fine detail.

5.3. ETH-80 Dataset

Table 2 lists the retrieval rate for each group and the average retrieval rate for the ETH-80 dataset for the proposed technique and a number of techniques in the literature [7]. A gray-scale version of the dataset is used. The results reflect the superiority of the proposed technique.

5.4. Logo Dataset

The proposed technique is also tested on a 500 logo dataset (see Appendix A for the full dataset). Some of the search

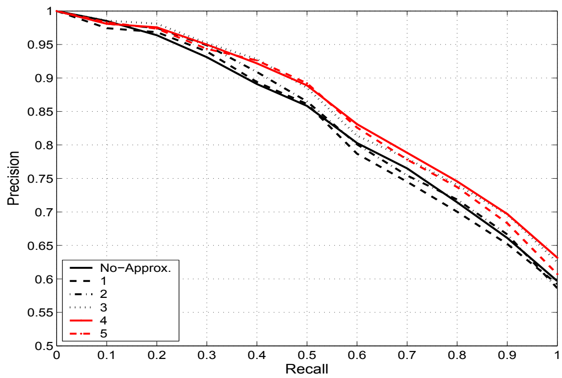


Fig. 6. Kimia-99 PR curves at different levels of approximation.

Table 2. ETH-80 retrieval rates for the proposed technique compared to that of some techniques in the literature [7].

PCA Binary	PCA Gray	Cont. Greedy	Cont. DynProg	**Proposed**
83.41%	82.99%	86.4%	86.4%	**88.26%**

results are available in Figure 7. The search logo is the first on top followed by the first four retrieved logos in a descending order, based on the cross-correlation to the search logo. The first match (the search image itself) is removed to eliminate redundancy. The results reflect the new technique's capability to reflect the perceptual similarity between objects.

6. CONCLUSION

In this paper, a new shape representation and description technique is introduced. The proposed technique utilizes the Hilbert space filling curve to scan the image creating a compact 1D representation of it. After the resulting vector is smoothed, it is sampled to produce a shape feature vector of the desired size. This linear time technique proves to be invariant to translation, scaling, both uniform and non-uniform. Experimental results are very promising

7. REFERENCES

[1] F., S. Abbasi, and J. Kittler. Efficient and robust retrieval by shape content through curvature scale space. In *International Workshop on Image Databases and MultiMedia Search*, pages 35–42, Amsterdam, The Netherlands, 1996.

[2] Serge Belongie, Jitendra Malik, and Jan Puzicha. Matching shapes. In *IEEE International Conference on Computer Vision*, pages 454–463, 2001.

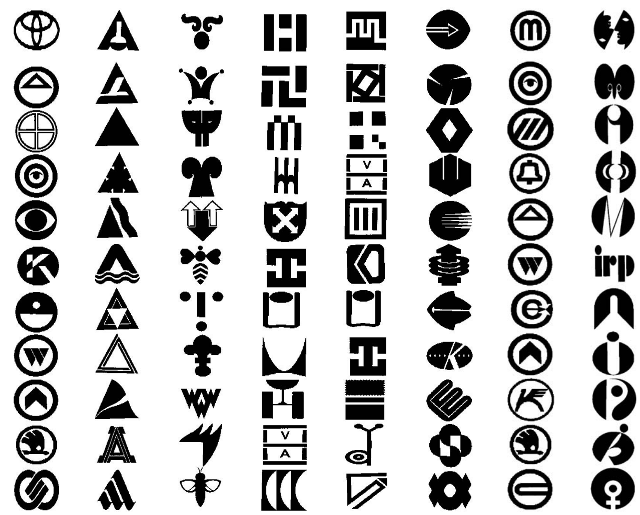


Fig. 7. Logo search results.

[3] L. J. Latecki and R. Lakamper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1185–1190, 2000.

[4] T.B. Sebastian, P.N. Klein, and B.B. Kimia. Recognition of shapes by editing shock graphs. In *Eighth IEEE International Conference on Computer Vision*, pages 755–762, 2001.

[5] G.J. Lu and A. Sajjanhar. Region-based shape representation and similarity measure suitable for content-basedimage retrieval. *Multimedia Systems*, 7(2):165–174, 1999.

[6] H. Kim and J. Kim. Region-based shape descriptor invariant to rotation, scale and translation. *Signal Processing: Image Communication*, 16:87–93, 2000.

[7] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *International Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 409–415, Madison, Wisconsin, June 2003.