A HIERARCHICAL CHARACTERIZATION SCHEME FOR IMAGE RETRIEVAL

Liu Wenyin
Microsoft Research, China
No. 49, Zhichun Road
Beijing 100080,P.R.China
E-mail: wyliu@*microsoft.com*

Tao Wang*
Dept. of Computer Sci. & Tech.
Tsinghua University
Beijing 100084, P.R.China
E-mail: wangtao72@yahoo.com

HongJiang Zhang
Microsoft Research, China
No. 49, Zhichun Road
Beijing 100080,P.R.China
E-mail: hjzhang@microsoft.com

ABSTRACT

In this paper, we present a study of hierarchically characterizing image content from coarse level to fine level conducted using a series of shape features as a case in point. In this work, we have empirically proved that this hierarchical characterization is effective and necessary. We have used this hierarchical multimedia content characterization scheme in our shape retrieval experiments and see that the hierarchical method obtains similar performance in retrieval similar shapes to that of the simultaneous method (calculating similarity based on all features at a time), while the efficiency has been increased significantly.

1. INTRODUCTION

Rapid developments of Internet and multimedia technologies have resulted in increasing amount of information in multimedia forms. Information retrieval based on multimedia content similarity is an effort to handle this "information explosion" problem. Content based image retrieval (*CBIR*) systems, e.g., QBIC [2] and Virage, are examples of such efforts. However, as machine understanding of image content is still at a much lower level than human intelligence, more efficient and more effective characterization of image content is still an open problem.

Current CBIR systems use many features, including texture, color, and shape to characterize image content. Similarity between two images is usually measured in terms of the differences between corresponding features of the two images using one of the many similarity models, as listed by Jain et al. [4]. These features are used simultaneously in the similarity assessment. That is, an overall similarity is calculated from all these feature values and used for the final similarity judgment between the two images. In the simultaneous method, usually all features are considered equal in characterizing the image content, though their weights in the similarity assessment may be different. Alternatively, we may assess the similarity by hierarchically filtering out those significantly dissimilar images using as less as possible features. Only those images that are sufficiently similar deserve detail assessment. In order words, an image needs further characterization only when higher-level characterization is not sufficient. In this case, we may say that the image content is characterized hierarchically by these features. Following this idea, we have proposed a scheme of hierarchical characterization of images focused on CBIR applications.

1.1. Why Hierarchical Characterization

Obviously, improving search efficiency is a major reason for using hierarchical characterization. The simultaneous method requires comparing all features of two images at one time, which is computationally expensive. However, hierarchical characterization method only extracts and use those features at necessary levels in the hierarchy, thus is more efficient than the simultaneous method. In the experiments shown lately in this paper, the top-level shape feature extraction uses only 1/67 of the time used for the low-level feature extraction. Another major reason is due to the curse of dimensionality [7], which means too many features may worsen classification. E.g., similar images that have identical values for a small number of relevant features may nevertheless be distant from one another in the feature space (due to other irrelevant features) and may therefore be judged by machines as dissimilar using the simultaneous method. Hence, the importance of different image features in image characterization or similarity assessment should not be equal. If we select the relevant or salient features at a proper hierarchical level and those irrelevant features to a very low level, and only check these low level features when necessary, we may avoid the curse of dimensionality. In other words, if we select the features correctly in building up an image characterization hierarchy, we may improve the performance of similarity assessment.

Finally, complexity of similarity models is another reason for hierarchical characterization. Jain et al. have summarized many similarity models [4]. However, none of them can precisely reveal the real natures of human perceptual model. Krumhansl proposed the distance-density model of similarity, in which similarity is assumed to be a function of both the feature difference and the

* The work reported in the paper was performed while this author was an intern at Microsoft Research, China, Feb.-Apr. 2000.

feature density of the feature space [5]. Since the feature density is also difficult to model, this similarity model is still difficult to apply. Since in general we know little about the density of the feature space and the property of the similarity model, the similarity measures are usually not accurate. However, we can access comparable, meaningful, and accurate similarity to a certain extent if we observe a small subspace around the point of interest (or anchor point referred to later) in the feature space. We can do so because the similarity function can be considered solely linear to the feature distance in such a small space. In other words, similarity assessment or comparison makes sense only when all the feature differences are smaller than some thresholds. Therefore, we can use the threshold to filter out those points far from the anchor point hierarchically using the available features, since they are considered significantly dissimilar, unnecessary to assess the total similarity. We can calculate the total similarity in detail for those surviving the filtering.

Admittedly, there is a risk of excluding strongly similar objects that are however quite far away from the anchor point. Hence, the last two reasons may deteriorate the characterization efficacy of the hierarchical method. We need to lower down this deterioration to a certain extent when building the characterization hierarchy.

Based on the above reasons, we propose in this paper a hierarchical scheme for image content characterization from coarse level to fine level as follows. Any image needs at least its top level features to be extracted for characterization. If such characterization is satisfied, the entire process of image characterization stops. Otherwise middle level features are extracted. If these features are still insufficient to characterize the image, lower features are further extracted. The entire characterization process is completed when available features are sufficient or no more features can be extracted.

We have used the hierarchical scheme in our shape retrieval system and have empirically proved that this hierarchical characterization is effective and necessary. The hierarchical method obtains similar accuracy in retrieval shapes to that of the simultaneous method but with a significant improvement in efficiency.

2. SHAPE CHARACTERIZATION BY HIERARCHICAL FEATURES

Among the many classes of visual features of objects or their images, shape features are intuitively clearer and easier to describe. Intuitive shape features include topological or geometric features while others include transform-based features. In this section, we present the proposed scheme for hierarchical characterization of images using shape features.

2.1 Shape Features

We refer to shape as a 2D closed contour and intend to represent a shape using a polygon. Hence, shape in this paper is a well defined concept and has a clear description. Shape analysis researches have been conducted for many years and quite a lot of shape features have been proposed to characterize shape and shape similarity [1]. Such features are usually independent of size and orientation, including eccentricity, compactness, and other moment-based features, etc.

We have selected the three global topological shape features, namely, eccentricity [6], compactness [3], and solidity, and a set of Normalized Fourier Descriptors (NFDs) based on the Fourier Descriptors (FDs) [10] to characterize shapes. All these feature values are within the range between 0 and 1. They are also invariant of translation, orientation, and scaling. They are therefore suitable to characterize shape uniformly.

In order to reduce noises and redundant data in the polygon form of the shape contour, we apply the method of Sklansky and Gonzalez [9] to remove more non-critical points from the polygon and approximate it using a more concise one. Treating the polygon as a periodic discrete signal, we define these features as follows.
1) *Eccentricity* [6] is defined in Eq (1).

$$Eccentricity = \frac{I_{\min}}{I_{\max}} = \frac{u_{20} + u_{02} - \sqrt{(u_{20} - u_{02})^2 + 4u_{11}^2}}{u_{20} + u_{02} + \sqrt{(u_{20} - u_{02})^2 + 4u_{11}^2}} \quad (1)$$

where, $u_{p,q} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$ is the (p, q) order central moments of the shape and can be calculated from the polygon vertexes using the efficient method of Leu (1991). As can be seen from (1), eccentricity is in fact the ratio of the short axis' length ($I_{\min}$) to the long axis' length ($I_{\max}$) of the best fitting ellipse of the shape.

2) *Compactness* is defined in this paper using the following equation based on the definition by Jain [3].

$$Compactness = \frac{4\pi A}{P^2} \quad (2)$$

where, P is the perimeter and A is the area of the polygon. Compactness expresses the extent to which a shape is a circle. A circle's compactness is 1 and a long bar's compactness is close to 0.

3) *Solidity* is defined in this paper as following,

$$Solidity = \frac{A}{H} \quad (3)$$

where A is the area of the polygon and H is the convex hull area of the polygon. Solidity describes the extent to which the shape is convex or concave.

4) *Normalized Fourier Descriptors*

Fourier descriptors (FD) [10] are the coefficients of the discrete Fourier transform, which are resulted from the frequency analysis of shape. FD plays an important role in shape characterization and discrimination [8]. Although they are independent of translation and orientation, they are actually dependent on the shape scale. We normalize them and make the normalized FDs (NFDs) scale independent. First of all, we parameterize the polygon to a parameter curve with N points by an equal arc length (N=256 is used in our work presented in this paper). We calculate the discrete Fourier transform of these N points expressed as {(x+iy)} using the following equation.

$$z(k) = F(x+iy) = \sum_{n=0}^{N-1} (x(n)+iy(n)) e^{-j\frac{2\pi}{N}kn}, \quad k=0,1,2\ldots N-1 \tag{4}$$

Assuming that a shape is scaled by r times, rotated by an angle φ, and shifted by an offset (x_0, y_0), we have the following deduction about the Fourier transform $z'(k)$ of the new shape.

$$\begin{aligned} z'(k) &= F[(x+iy)re^{j\varphi} + (x_0+iy_0)] \\ &= re^{j\varphi}F(x+iy) + F(x_0+iy_0) = re^{j\varphi}z(k) + F(x_0+iy_0) \\ &\Rightarrow z'(0) = re^{j\varphi}z(0) + C \\ & z'(k) = re^{j\varphi}z(k) \qquad k=1,2,\ldots N-1 \\ &\Rightarrow |z'(k)| = r\,|z(k)| \qquad k=1,2,\ldots N-1 \end{aligned}$$

$$\Rightarrow \frac{|z'(k)|}{|z'(1)|} = \frac{r\,|z(k)|}{r\,|z(1)|} = \frac{|z(k)|}{|z(1)|} \qquad k=1,2,\ldots N-1 \tag{5}$$

where, $F(x_0+iy_0)$ is a Fourier transform of a constant. It is a pulse C at k=0 and is zero when k is non zero.

As can be seen from the above deduction, the Fourier transform of the shape changes its phase when shape is rotated and changes its amplitude when it is scaled and $z(0)$ is changed when it is shifted. According to (5), we defined the normalized FD $d(k)$, as follows.

$$d(k) = \frac{|z(k)|}{|z(1)|} \qquad k = 2,3,\ldots,N-1 \tag{6}$$

NFD $d(k)$ is invariant of translation, orientation, and scale.

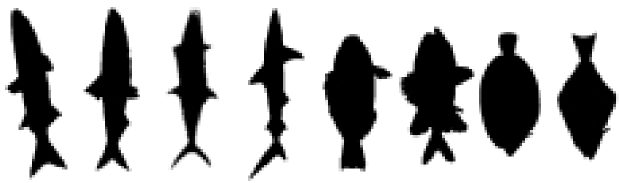
2.2 Hierarchical Characterization of Shape

Following the idea presented in Section 1, shape can be characterized hierarchically by those features discussed in Section 2.1. For example, some features, e.g., eccentricity and compactness, are global topological features to characterize shape in the overall sense, while other features, e.g., Fourier Descriptors (FDs), are local geometric features to characterize details of the shape. Chen has proposed that global topological properties and local geometric properties belong to different hierarchical levels of perception [1]. Obviously, global features deserve a higher level than local features in the hierarchy. Although two shapes may have similar local features, they may not be considered similar if their global features are not similar. Hence, we build the shape characterization hierarchy of three levels accordingly. The three global topological features (eccentricity, compactness, and solidity) as a group are at top of the hierarchy. We use the 11 lower frequency NFDs ($d(k)$, k=2,3, …, 12) at the middle level and use other high frequency NFDs at the lowest level. In fact, we found that the 11 lower frequency NFDs at the middle level are able to characterize the shape to the very details for our shape retrieval experiment. So in our experiment, we only use two level of characterization. Another reason of not using the high frequency NFDs is that they are not stable in noisy situations and they all have very small (undiscriminating) values. We use the similarity model of city-block distance to assess shape similarity at each level when similarity assessment is necessary.

3. THE SHAPE RETRIEVAL EXPERIMENT

We have selected 8 different kinds fish images, as shown in Figure 1, to build the synthesized database. From each of the 8 images we create 11 new images by rotating, translating, and zooming it randomly. Adding another 4 images, we make the size of our synthesized image database 100. The shape based retrieval results using different features and schemes are listed in Table 1. In Table 1, GF means the retrieval is performed using only the global features, FD means the retrieval is performed using only the NFD features, G&F means using both global features and NFD features simultaneously, and HG&F means using the hierarchical method with the global feature similarity threshold of 0.18. From Table 1 we can see that the global features and the NFD features do not work well individually in discriminating shapes. The G&F method, however, works well, being able to recognize all the same kind of fish perfectly, with a recall rate of 100% (or 0% false alarm rate). The hierarchical method has the same retrieval performance with the G&F method, but is much faster than the G&F method. The result shows that the retrieval of shape images using these features hierarchically is effective for both retrieval accuracy and efficiency.

Another experiment is conducted in a larger fish database. There are 1100 fish images/shapes in our fish image database. Features are preliminarily extracted off-line using an average time 0.54ms per image in the extraction of global features and 36.37ms per image for the extraction of the NFD features. Similarity is on-line calculated. The retrieval time is 10 ms (on a PII400 machine) by the GF method, 10 ms by the FD method, 20 ms by the G&F method, and 11 ms by the hierarchical (HG&F) method. Therefore, that hierarchical retrieval method is almost twice faster than the normal simultaneous G&F method while preserving the same performance. The retrieval results by different methods are shown in Figure 2. The top left one in each pane of Figure 2 is the query shape. As we can see from Figure 2, the GF method has successfully found those shapes that have similar bounding boxes but may have different shape details. The FD method has found those shapes that are similar in their shape details but may have different global features. The shapes found by the G&F method found are almost similar both globally and in shape detail. The HG&F has found as many shapes as the G&F method has, but is faster than the G&F method.

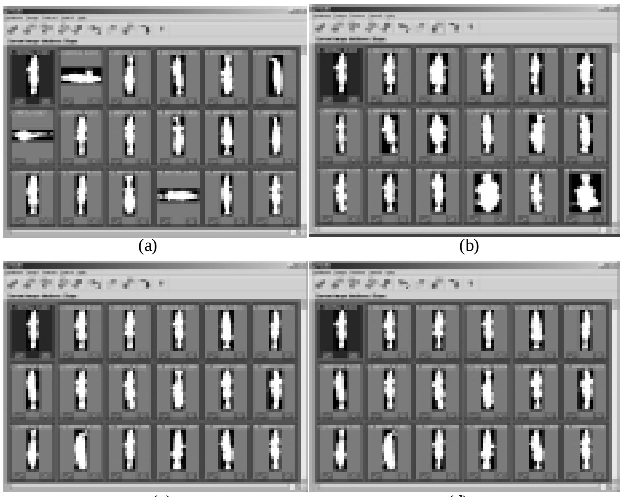


F1 F2 F3 F4 F5 F6 F7 F8

Figure 1: The 8 shape seeds of the synthesized database.

Table 1. Shape based retrieval of the shapes in Figure 1.

Method	F1	F2	F3	F4	F5	F6	F7	F8	Time(ms)
GF	86.1%	91.7%	98.6%	100%	73.6%	100%	100%	43.1%	0.10
FD	98.6%	100%	100%	100%	100%	100%	92.4%	87.8%	0.10
G&F	100%	100%	100%	100%	100%	100%	100%	100%	0.20
HG&F	100%	100%	100%	100%	100%	100%	100%	100%	0.11



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

Figure 2. Shape retrieval result by (a) GF, (b) FD, (c) G&F, (d) HG&F.

4. CONCLUSION REMARKS

In this paper, we have presented our study on the effectiveness of hierarchically characterizing image content from coarse level to fine level using a series of shape features as a case in point. Empirically we have proved that the proposed hierarchical characterization scheme is more efficient than the simultaneous method, while achieving the same retrieval accuracy. In the shape retrieval experiment with shape features preliminarily extracted off-line, similarity assessment in the hierarchical method is twice faster than in the simultaneous method. If we take into account the online feature extraction time (e.g., as required in an on-line image search on the web), which is 67 times faster for the three global features than for the local NFD features, we believe the hierarchical method will be much faster than the simultaneous method.

REFERENCES

[1] Chen L, "Topological structure in visual perception", *Science* 218(12):699-700, 1982.

[2] Flickner M et al. "Query by Image and Video Content", *IEEE Computer* 28(9):23-32, 1995.

[3] Jain AK, *Fundamentals of digital image processing*, Prentice-Hall, 1989.

[4] Jain R, Murthy SNJ, and Chen PLJ, "Similarity Measures for Image Databases", *Proc. IEEE Conf. On Fuzzy Systems*, vol. 3, pp. 1247-1254, 1995.

[5] Krumhansl CL, "Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density", *Psychological Review* 85:445-464, 1978.

[6] Leu J-G, "Computing a shape moments from its boundary", *Pattern Recognition* 24:949-957, 1991.

[7] Trunk GV, "A Problem of Dimensionality: A Simple Example", *IEEE Trans. On PAMI* 1:306-307, 1979.

[8] Persoon E and Fu KS, "Shape Discrimination using Fourier Descriptors", *IEEE Trans. on SMC* 7:170-179, 1979.

[9] Sklansky J and Gonzalez V, "Fast Polygonal Approximation of Digitized Curves", *Pattern Recognition* 12:327-331, 1980.

[10] Zahn CT and Roskies RZ, "Fourier Descriptors for plane closed Curves", *IEEE Trans. On Computers* 21:269-28, 1972.