MPEG-7 Visual Shape Descriptors

Miroslaw Bober

***Abstract*—This paper describes techniques and tools for shape representation and matching, developed in the context of MPEG-7 standardization. The application domains for each descriptor are considered, and the contour-based shape descriptor is presented in some detail. Example applications are also shown.**

I. Introduction

OBJECT SHAPE features provide a powerful clue to object identity and functionality, and can even be used for object recognition. Humans can recognize characteristic objects solely from their shapes—proof that shape often carries semantic information. This distinguishes shape from other elementary visual features, such as color, motion, or texture, which, while equally important, usually do not reveal object identity. A large body of research has been devoted to shape-based recognition, retrieval, and indexing [1].

Many applications, including those concerned with visual objects retrieval or indexing, are likely to use shape features. For example, an intelligent video security system may use shape techniques to determine the identity of the intruder. Shape descriptors can be used in e-commerce, where it is difficult to use a text index to specify the required shape of the object, and query-by-example is simpler and faster. Naturally, the best retrieval performance can be achieved by using a combination of visual (and possibly other) feature descriptors, but that subject is outside the scope of this paper.

In the following section, an overview of the MPEG-7 shape descriptors is presented. Their functionality, properties, and application domains are briefly analyzed. The contour-based shape descriptor is then considered in detail and some examples of applications are given.

II. Overview of Shape Descriptors

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 [2], [4]. 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 Fig. 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*.

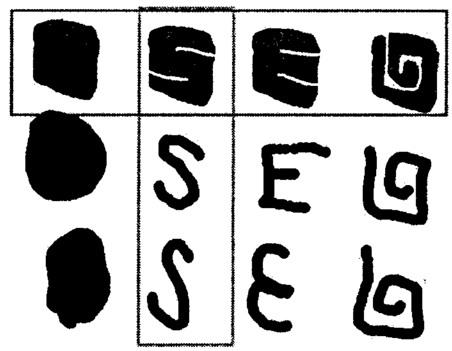

Fig. 1. Examples of contour- and region-based shape similarity.

Sometimes a 3-D model of the object is not known, but the application requires 3-D information. For example, in an application concerned with e-shopping, the user may have a mental picture of the product, e.g., a flower vase, but not the actual 3-D model from which 3-D descriptor could be extracted. Such a mental picture consists of a set of 2-D views, which when combined describe the 3-D properties of the object. To cater for such situations, a *2-D/3-D shape descriptor* has been developed.

The MPEG-7 shape descriptors were selected in extensive tests [3], where various properties of the descriptors were assessed. The technologies chosen for the standard are the ones that best fulfil the requirements defined by the experts. A good descriptor captures characteristic shape features in a concise manner, which renders itself to fast search and browsing. It is not required that the original shape can be reconstructed from the descriptor, and therefore a significant compactness can be achieved. It should be invariant to scaling, rotation, translation and to various types of shape distortions. The distortions may be a result of the imaging conditions; e.g., a 2-D shape may undergo a perspective transformation as a results of change in viewing angle. Some objects are flexible and the technique should be able to cope with nonrigid deformations. Finally, it should cope with typical distortions, which may result from the segmentation or extraction processes. We will now briefly introduce each shape descriptor, namely the 3-D shape descriptor, region-based shape descriptor, contour-based shape descriptor, and 2-D/3-D shape descriptor.

A. Shape Spectrum—3-D Shape Descriptor

The MPEG-7 3-D shape descriptor is based on the shape spectrum concept. It is an extension of the shape index, used previously as a local measure of 3-D shape [4] to 3-D meshes.

Manuscript received September 22, 2000; revised March 20, 2001.

The author is with the Visual Information Laboratory, Mitsubishi Electric Information Centre Europe, Surrey Research Park, Guildford GU2 7YD, U.K.

Publisher Item Identifier S 1051-8215(01)04988-6.

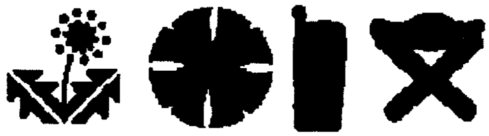

Fig. 2. Examples of shapes that can be efficiently described by the region-based descriptor.

The shape index of an orientable 3-D surface Σ, at point p, is defined as

$$\mathrm{SI}_{\Sigma} = 1/2 - 1/\pi \arctan \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}$$

where $k_1(p)$ and $k_2(p)$ denote the principal curvatures at point p; $k_1(p) \leq k_2(p)$. $\mathrm{SI}_{\Sigma}(p) \in [0, 1]$, except for planar regions, where the shape index is undefined. The shape index captures information about the local convexity of a 3-D surface.

Shape spectrum is defined as the histogram of the shape index, computed over the entire 3-D surface. For 3-D meshes, the shape index is computed for each vertex of the mesh. It is invariant to scaling and to Euclidean transformations.

By default, the descriptor uses histogram with 100 bins and each bin is represented by 12 bits. Two additional variables are used to form the descriptor. The first one expresses the relative area of planar surface regions of the mesh, with respect to the entire area of the 3-D mesh. The second one is the relative area of all polygonal components where reliable estimation of the shape index is not possible, with respect to the entire area of the 3-D mesh.

B. Angular Radial Transformation (ART)—Region-Based Shape Descriptor

The region-based shape descriptor expresses pixel distribution within a 2-D object region; it can describe complex objects consisting of multiple disconnected regions as well as simple objects with or without holes (Fig. 2).

Some important features of this descriptor are the following.

1) It gives a compact and efficient way of describing properties of multiple disjoint regions simultaneously.
2) Sometimes during the process of segmentation, an object may be split into disconnected sub-regions. Such an object can still be retrieved, provided that the information which regions it was split into is retained and used during the descriptor extraction.
3) The descriptor is robust to segmentation noise, e.g., salt and pepper noise.

The region-based shape descriptor belongs to the broad class of shape analysis techniques based on moments [7]. It uses a complex 2-D ART [2], [6], defined on a unit disk in polar coordinates.

From each shape, a set of ART coefficients F_{nm} is extracted, using the following formula:

$$\begin{aligned} F_{nm} &= \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle \\ &= \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta), f(\rho, \theta) \rho \, d\rho \, d\theta \end{aligned}$$

where $f(\rho, \theta)$ is an image intensity function in polar coordinates and $V_{nm}(\rho, \theta)$ is the ART basis function of order n and m. The basis functions are separable along the angular and radial directions, and are defined as follows:

$$\begin{aligned} V_{nm}(\rho, \theta) &= \frac{1}{2\pi} \exp(jm\theta) R_n(\rho), \\ R_n(\rho) &= \begin{cases} 1, & n = 0 \\ 2\cos(\pi n \rho), & n \neq 0. \end{cases} \end{aligned}$$

The default region-based shape descriptor has 140 bits. It uses 35 coefficients $(n = 10, m = 10)$ quantized to 4 bits/coefficient.

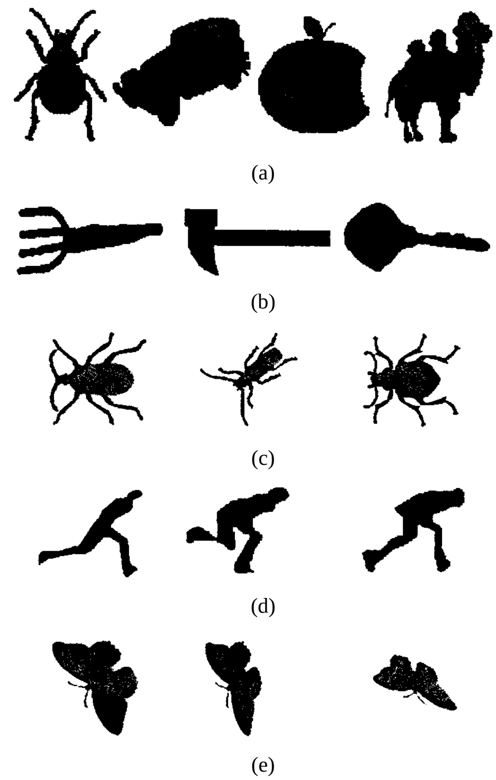

Fig. 3. Examples of shapes where a contour-based descriptor is applicable.

C. Contour-Based Shape Descriptor

The contour-based descriptor expresses shape properties of the object outline (contour). Objects for which characteristic shape features are contained in the contour are described efficiently by this descriptor [see Fig. 3(a)].

If a complex object consists of multiple disjoint regions, each region of component contours can be described separately, using the contour-based descriptor and an MPEG-7 description scheme. The descriptor, its properties, and example applications are considered in detail in Section III.

D. 2-D/3-D Descriptor

The 2-D/3-D descriptor can be used to combine 2-D descriptors representing a visual feature of a 3-D object seen from different view angles [10]. The descriptor forms a complete 3-D view-based representation of the object. Any 2-D visual descriptor, such as contour shape, region shape, color, or texture can be used. The 2-D/3-D descriptor supports integration of the 2-D descriptors used in the image plane to describe features of the 3-D (real world) objects. Experiments with 2-D/3-D descriptor and contour-based shape descriptor shown good performance in multiview description of the 3-D shapes.

III. Contour-Based Shape Descriptor

The contour-based shape descriptor is based on the curvature scale-space (CSS) representation of the contour [9], [8]. This representation has been successfully used for search and retrieval in the past, and has been further extended and optimized during the MPEG development phase. The key modifications include: 1) addition of global shape parameters; 2) transformation of the feature vector in the parameter space improving retrieval performance; and 3) a new quantization scheme supporting a compact representation of the descriptor. The descriptor has been selected in perhaps the most comprehensive comparative testing of performance of shape-based techniques to date, performed within MPEG. Techniques tested included Fourier-based techniques, shape representations based on Zernike moments, turning angles, and wavelet-based techniques.

A. Properties of the Contour-Based Shape Descriptor

The shape properties of contours are important for retrieval of semantically similar objects [1]. The descriptor is also very efficient in applications where high variability in the shape is expected, due to, e.g., deformations in the object (rigid or nonrigid) or perspective deformations. The description is robust to noise present in the contour.

Below, we list some of the important features.

1) It can distinguish between shapes that have similar region-shape properties but different contour-shape properties. Objects shown in Fig. 3(b) have a similar region pixel distribution, but different contour properties, and thus contour-based descriptor can efficiently differentiate between such shapes.
2) It supports search for shapes that are semantically similar for humans, even when significant intra-class variability exists, such as in the images of beetles [see Fig. 3(c)].
3) It is robust to significant nonrigid deformations [see Fig. 3(d)].
4) It is robust to distortions in the contour due to perspective transformations, which are common in the images and video [see Fig. 3(e)].

B. Descriptor Syntax and Extraction

The descriptor consists of the eccentricity and circularity values of the original and filtered contour (each 6 bits), the index indicating the number of peaks in the CSS image (6 bits), the height of the highest peak (7 bits) and the x and y positions on the remaining peaks (9 bits/peak) [2]. The average size of the descriptor is 112 bits/contour.

To create a CSS description of a contour shape, N equi-distant points are selected on the contour, starting from an arbitrary point on the contour and following the contour clockwise. The x coordinates of the selected N points are grouped together, and the y coordinates are also grouped together into two series X and Y. The contour is then gradually smoothed by repetitive application of a low-pass filter with the kernel (0.25, 0.5, 0.25) to X and Y coordinates of the selected N contour points. As a result of the smoothing, the contour evolves and its concave parts gradually flatten-out, until the contour becomes convex. A so-called CSS image can be associated with the contour evolution process (the CSS image does not have to be explicitly extracted, but is useful to illustrate the CSS representation). The CSS image horizontal coordinates correspond to the indices of the contour points selected to represent the contour $(1, \ldots, N)$, and CSS image vertical coordinates correspond to the amount of filtering applied, defined as the number of passes of the filter. Each horizontal line in the CSS image corresponds to the smoothed contour resulting from k passes of the filter. For each smoothed contour, the zero-crossings of its curvature function are computed. Curvature zero-crossing points separate concave and convex parts of the contour. Each zero-crossing is marked on the horizontal line corresponding to the smoothed contour and at the location corresponding to the position of this zero-crossing along the contour. The CSS image has characteristic peaks. The coordinate values of the prominent peaks $(x\_\mathrm{css}, y\_\mathrm{css})$ in the CSS image are extracted. The peaks are ordered based on decreasing values y_css, transformed using a nonlinear transformation and quantised. In addition, the eccentricity and circularity of the contour are also calculated, quantized, and stored.

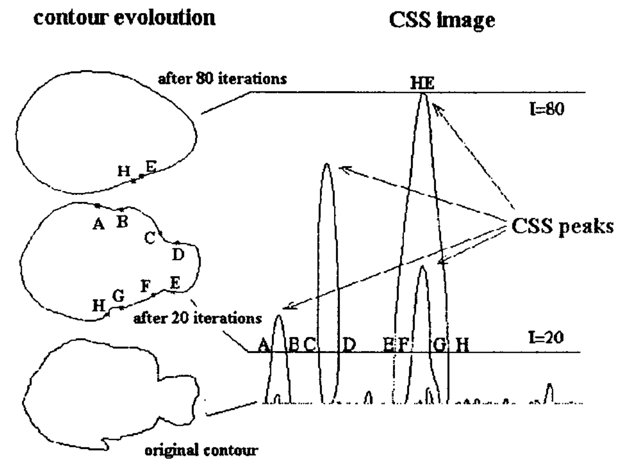

Fig. 4. Example of contour evolution and corresponding CSS.

Fig. 4 shows the shape evolution during filtering (left) and the corresponding CSS image. The contour curvature zero crossings $(A, B, \ldots H)$ and the corresponding points on the CSS image are marked.

C. Example Application—Video Browsing

Fig. 5 shows an example of a Web-based video clip search application using MPEG-7 contour shape and dominant color descriptors. A cartoon character is selected from the group of images in the upper part of the display. The contour shape descriptor and dominant color are extracted from the selected object or required shape can be drawn directly by the user. A search is performed on the descriptions stored on the servers and key frames from the video clips containing similar characters are displayed to the user (lower row of images). The user can then select the required clip, which is transmitted over the network and played back.

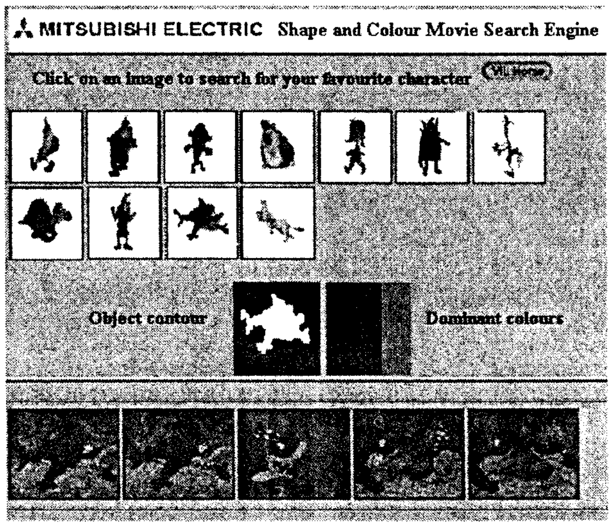

Fig. 5. Web-based video clip search application using MPEG-7 contour shape and dominant color descriptors.

IV. CONCLUSION

A set of versatile shape descriptors has been developed, supporting an entire spectrum of possible applications. Their performance has been tested, and it has been shown that the descriptors are efficient, concise, and easy to extract and match.

ACKNOWLEDGMENT

The author would like to thank all participants in the MPEG "Shape Experimentation Group," especially W.-Y. Kim, W.-S. Kim, K. Muller, J. Ohm, J. Cooper, H.-K. Kim, J. D. Kim, T. Zaharia, and F. Preteux for their technical contributions toward the development of the MPEG shape descriptors.

REFERENCES

[1] S. Ullman, *High Level Vision*. Cambridge, MA: MIT Press, 1997.
[2] *The MPEG-7 Visual Part of the XM 4.0*, ISO/IEC MPEG99/W3068, Dec. 1999.
[3] S. Jeannin and M. Bober, "Description of core experiment for motion/shape," ISO/IEC N2690, Seoul, Korea, Mar. 1999.
[4] T. Zaharia, F. Preteux, and M. Preda, "The 3D shape spectrum descriptor," ISO/IEC MPEG/M5242, Melbourne, Australia, Oct. 1999.
[5] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image Vis. Comput.*, vol. 10, no. 8, 1992.
[6] W.-Y. Kim and Y.-S. Kim, "A New Region-Based Shape Descriptor," ISO/IEC MPEG99/M5472, Maui, Hawaii, Dec. 1999.
[7] C.-C. Ten and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 496–513, July 1988.
[8] M. Bober, "Shape descriptor based on curvature scale space,", Lancaster, U.K., MPEG-7 proposal P320, Feb. 1998.
[9] F. Mokhtarian and A. K. Mackworth, "A theory of multiscale, curvature-based shape representation for planar curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 789–805, Aug. 1992.
[10] K. Müller, J.-R. Ohm, J. Cooper, and M. Bober, "Results of 2D/3D shape core experiments MS-4," ISO/IEC MPEG99/M6190, Beijing, China, July 2000.