AUTOMATED REGION SEGMENTATION USING ATTRACTION-BASED GROUPING IN SPATIAL-COLOR-TEXTURE SPACE

Yong Rui, Alfred C. She and Thomas S. Huang

Department of Electrical and Computer Engineering, Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
E-mail:yrui@ifp.uiuc.edu, a-she@uiuc.edu

ABSTRACT

Recently much attention has been paid to Image Content-Based Retrieval Systems (CBRS). One important goal in CBRS is to extract *local* low level image features such as color, texture and shape, to allow queries based on these features.

A large CBRS containing tens of thousands of images requires an automatic feature-extraction method since human aided segmentation is impractical. We address this problem in a particular application setting by using an *attraction* based grouping method in Spatial-Color-Texture space. The attraction concept makes this approach similar to human aided segmentation. Experimental results show that the method is reasonably better than existing methods, and has the potential to be used in other CBRS-related applications.

1. INTRODUCTION

Image content-based retrieval systems (CBRS) have recently been gaining attention [1-3]. The low-level image features currently used in CBRS include color, texture, and shape. Global image features such as the global color histogram, global texture features (coarseness, contrast, and directionality), and so on [1,4], are easily computed. However, local features are needed to support more specific queries (e.g., "Find images that contain two regions that are green and have coarse texture.") and layout queries (e.g., "Find images that contain a small red region surrounded by a large blue region."). Region segmentation is the basis for extracting local color and texture features. Obviously region segmentation is used in shape-related queries as well [5].

This work was supported by the NSF/DARPA/NASA Digital Library Initiative under Cooperative Agreement No. 94-11318.

It is well known that automatic segmentation of images in general is an unresolved issue. We address this issue in a particular application setting, and discuss how to extend the proposed method to other applications.

Generally speaking, spatial, color, or texture features alone do not have enough discriminating power to properly segment out image regions. A hybrid Spatial-Color-Texture space is defined in section 2. Section 3 describes c-means clustering in Color-Texture space. Section 4 describes in detail the procedure for attraction-based grouping. Experimental results and conclusions are in sections 5 and 6.

2. SPATIAL-COLOR-TEXTURE SPACE

2.1 Spatial space

Spatial space is just the 2-D Cartesian space spanned by the x and y image coordinates.

2.2 Color space

Although the RGB color space is easy to understand and work with, it is not perceptually uniform. Other color spaces have been proposed, such as HSV, CIE-LAB, CIE-LUV, and Munsell [6,7]. Roughly speaking, a common property of these color spaces is that they separate color space into perceptually independent components. This allows us to work with one component without significantly affecting the other two.

Furthermore, the HSV, CIE-LAB, and Munsell color spaces also attempt to make the color space perceptually uniform. The Euclidean distance in these spaces is close to the perceived color distance under certain viewing conditions.

We chose the HSV (hue, saturation, and value) color space for simplicity.

2.3 Texture space Although many sophisticated texture features have been proposed [8], a simple yet efficient feature, *edge density*, is used in this paper. We chose this feature for two reasons. First, if this simple feature works well, choosing better feature will achieve even better results. Therefore this feature serves as a worst case example. Second, this feature is computationally cheap, which is attractive to a large-size CBRS.

The procedure used in this paper to obtain the edge density is the following:

1. For each pixel, find the horizontal and vertical edge magnitudes $G_x[i][j]$ and $G_y[i][j]$ using the Sobel Gradient Operator;

2. For each pixel, compute the edge magnitude

$$G[i][j] = |G_x[i][j]| + |G_y[i][j]|$$

3. For each pixel, compute the edge density $Dns[i][j]$ within a $(2m+1) \times (2m+1)$ window

$$Dns[i][j] = \frac{1}{(2m+1)^2} \sum_{k=i-m}^{k=i+m} \sum_{l=j-m}^{l=j+m} G[k][l]$$

In step 2 we use the absolute value to increase computational speed.

The Spatial-Color-Texture space is a 6-dimensional space. C-means clustering is done in the 4-D Color-Texture sub-space, while attraction-based grouping is done in the full 6-D space.

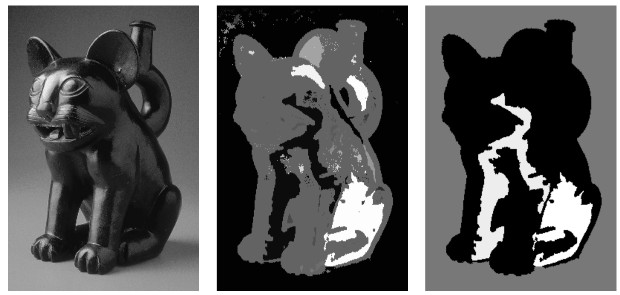
3. C-MEANS CLUSTERING IN COLOR-TEXTURE SUB-SPACE

For a typical natural image, there is a high number of different colors and textures. C-means clustering is one way to reduce the complexity while retaining salient color and texture features.

1. Randomly pick c starting points in the Color-Texture space as the initial means.
2. Cluster each point as belonging to the nearest neighbor mean.
3. Compute the new mean for each cluster.
4. Repeat 2 and 3 until all the clusters converge (i.e. when the number of pixels and mean value of each cluster does not change).

After this procedure, we have c clusters, each of which may corresponds to a set of image pixels. We define *cluster* as a natural group which has similar features of interest [9]. The image pixels corresponding to a particular cluster may or *may not* be spatially contiguous. We define a *region* as one of the spatially connected regions corresponding to a cluster.

The c-means clustering generally produces regions of various sizes; some of the regions are very small (containing only a few pixels). We consider these regions as speckle noise (see Figure 1b) and set a minimum region size threshold to filter out these small regions. The deleted regions are merged with the largest neighboring region (Figure 1c).



(a) (b) (c)

Figure 1: (a) Original color image; (b) After clustering in Color-Texture space; (c) After thresholding. Note: Different intensities in Figures 1(b-c), 2(a-c), and 5b indicate different regions and are not related to the original image intensity.

In general, color or texture alone does not have enough discriminating power to properly segment an image. If the clustering is done in color space, the background regions can grow into the object regions due to shading or similar background colors. In Figure 2a, the animal's mouth and part of its ear are eroded by the upper background region.

(a) (b) (c)

Figure 2: (a) Clustering in color space only; (b) Clustering in texture space only; (c) Clustering in Color-Texture space.

Similarly, clustering in texture space only will not produce a correct segmentation either. In Figure 2b, the smooth boundary between the object and background are completely destroyed. However, when the clustering is done in Color-Texture space, we obtain a good segmentation (Figure 2c).

4. ATTRACTION-BASED GROUPING

Grouping in this paper refers to the process of grouping several regions into a shape.

After c-means clustering we have c clusters, each corresponding to several spatial regions. The next step is to extract the desired object from the regions.

Normally the clustered image will not have a uniform background as in Figure 2c, so we need to a method that can deal with non-uniform backgrounds.

One way to do this is to define a threshold in Color-Texture space. If a region's Color-Texture feature is above the threshold, then this region is considered as the object; otherwise, considered as the background. One obvious disadvantage of this thresholding method is that the threshold is image-dependent. We propose an attraction based grouping method (ABGM) to overcome this disadvantage. The method is motivated by the way the human visual system might do the grouping.

As defined in physics,

$$F_{12} = G\frac{M_1 M_2}{d^2}$$

reflects how large the attraction is between the two masses M_1 and M_2 when they are of distance d. In ABGM, we use the similar concept, but now M_1 and M_2 are the size of the two regions, and d is the Euclidean distance between the two regions in 6-D Spatial-Color-Texture space.

The ABGM method is described as follows:

1. Choose attractor region A_i's from the clustered regions according to the knowledge of the application at hand.
2. Randomly choose an unlabeled region R_j. Find the attractions F_{ij} between A_i and R_j.
3. Associate region R_j with the attractor A_i that has the largest attraction to R_j.
4. Repeat steps 2 and 3 until all the regions are labeled.
5. Form the output segmentation by choosing the attractor of interest and its associated regions.

Note that if the attractor is bigger or closer (in 6-D space) to a unlabeled region, its attraction will be larger and thus the unlabeled region will be labeled to this attractor with higher probability. This is what a human visual system might do in the labeling process.

5. EXPERIMENTAL RESULTS

The flowchart in Figure 3 shows the complete ABGM method whose components were described in Sections 3 and 4.

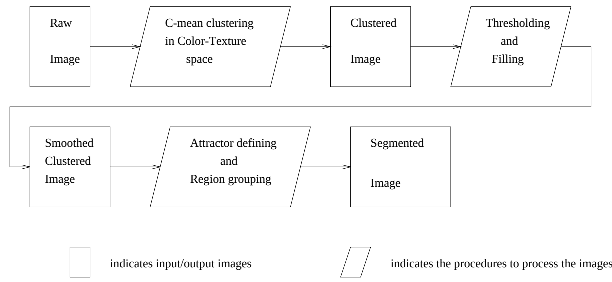


Figure 3: Flowchart of the ABGM.

The output segmentation depends on how the attractors are defined, i.e. the number and locations of the attractors. The number of the attractors defines how many shapes are of interest, while the locations of the attractors specify the initial features of the attractors.

The procedure shown in Figure 3 is a general procedure which can be specialized to work with different applications. We applied the ABGM to our multimedia database prototype – MARS (Multi-media Analysis and Retrieval System). MARS is designed to support both global and local queries on color, texture and shape. Currently there are 500 images in the prototype's database.

The ABGM is used in two subsystems in MARS. One is the shape-related query subsystem, the other is the local color/texture query subsystem.

5.1 Shape-related query subsystem

The purpose of using ABGM in this subsystem is to correctly segment out the shapes such that MARS can support shape-based queries.

The images in our prototype normally have one main object and normally the object is not on the boundary. Having this application knowledge, we define five attractors, with four on each corner being the background attractors and one in the center of the image being the object attractors.

The experimental results are promising. Out of the 500 images, over 70 percent can be correctly segmented. Figure 4c is a typical segmentation. Again, as mentioned in section 3, color or texture feature alone can not provide good segmentation (see Figures 4(a, b)).

Even when the ABGM is done in the 6-D Spatial-Color-Texture space, 30 percent of the images are incorrectly segmented. However, most of the incorrect segmentation results are caused by the multiple objects in the image whereas we define only one object attractor.

5.2 Local color/texture query subsystem

(a) (b) (c)

Figure 4: (a) Grouping in Spatial-Color space; (b) Grouping in Spatial-Texture space; (c) Grouping in Spatial-Color-Texture space.

In addition to shape-based queries, the ABGM can readily support local color, texture, and layout queries. The smoothed clustered image (Figure 3) contains all the information needed – the size and location, color, texture, and shape of each region.

A typical smoothed clustered image is given in Figure 5b. We can see that the regions containing salient color/texture features are properly segmented (e.g. umbrella, cart, palm trees, people, etc.). This region information allows the CBRS to support local feature queries. We are currently implementing this subsystem.

(a) (b)

Figure 5: (a) The original multi-shape color image; (b) Smoothed clustered image.

6. CONCLUSION

A new method of automated shape segmentation is proposed in this paper. The main features of this method are:

1. The concept of *attraction* makes this method behave in a way similar to the way a human might do the grouping process.
2. Since the ABGM is *automated*, it is useful in large CBRS.
3. The clustering and grouping in the *hybrid* Spatial-Color-Texture space has a relatively high discriminating power.

7. REFERENCES

1. C. Faloutsos, etal. Efficient and Effective Querying By Image Content, IBM Research Report, Aug. 3, 1993.
2. John R. Smith and Shih-Fu Chang, Tools and Techniques for Color Image Retrieval, IS & T/SPIE proceedings Vol. 2670, Storage & Retrieval for Image and Video Databases IV.
3. H. J. Zhang, etal. Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution, Submitted to ACM Multimedia, San Francisco, Nov. 5-9, 1995.
4. Hideyuki Tamura, etal. Texture Features Corresponding to Visual Perception, IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-8, No. 6, June 1978.
5. Yong Rui, Alfred C. She and Thomas S. Huang, Modified Fourier Descriptor for Shape Representation – A Practical Approach, submitted to First Int'l Workshop on Image Databases and Multi Media Search, Amsterdam, The Netherlands, 1996.
6. John R. Smith and Shih-Fu Chang, Single Color Extraction and Image Query, ICIP 1995.
7. Makoto Miyahara, etal. Mathematical Transform of (R,G,B) Color Data to Munsell (H,V,C) Color Data, 650 / SPIE Vol. 1001 Visual Communications and Image Processing 1988.
8. Philippe P. Ohanaian and Richard C. Dubes, Performance Evaluation for Four Classes of Textural Features, Pattern Recognition, Vol. 25, No. 8, pp. 819-833, 1992.
9. Richard O. Duda, etal. Pattern Classification and Scene Analysis, Wiley-Interscience Publication.