AUTOMATIC THRESHOLD DETERMINATION OF CENTROID-LINKAGE REGION GROWING BY MPEG-7 DOMINANT COLOR DESCRIPTORS

Fatih Porikli

Mitsubishi Electric Research Labs
Murray Hill, NJ 07974, USA

ABSTRACT

The emerging MPEG-7 standard embodies a visual descriptor that will be associated with the dominant colors of an image. In this contribution, a threshold adaptation method for region based image and video segmentation that takes the advantage of the MPEG-7 dominant color descriptor is presented. This method enables assignment of region growing parameters without any low-level processing. In the standard, the dominant colors proposed to be extracted by clustering of color histograms. This property is used to determine color homogeneity that is formulated into Lorentzian-based color distance norm and corresponding thresholds. The proposed algorithm is compared with other region growing algorithms, and results show that the threshold adaptation performs faster and more robust.

1. INTRODUCTION

The idea of region growing is one of the most fundamental and well known concepts used in image and video segmentation. Various research has been done in segmentation of images and video sequences using region growing techniques. Some of this work utilize experimentally set color distance thresholds [3], iteratively relaxing thresholds [4], navigation into higher dimensions to solve a distance metric formulation with user set thresholds [5], hierarchical connected components analysis with predetermined color distance thresholds [6], etc. In region growing, individual pixels that satisfy some neighborhood constraint are merged if their attributes, such as color and texture, are similar enough. Similarity can be established by applying a local or global homogeneity criterion. Usually, homogeneity criterion is implemented in terms of a distance function and corresponding thresholds. It is the formulation of the distance function and its thresholds that has the most significant effect on the segmentation results of a region growing method. Most methods either use predetermined color thresholds for every image, or mine for thresholds to adapt them to the input. Threshold adaptation may involve considerable amount of computation.

Meanwhile, MPEG-7 will be a standardized description of various types of multimedia information [1]. This description will be associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. The standard does not comprise the extraction of descriptions but ways to define other descriptors as well as structures. Audio and video material that has MPEG-7 data associated with it, can be indexed and searched for. This material may include: still pictures, graphics, 3D models, audio, speech, video, and information about how these elements are combined in a multimedia presentation. One of the descriptors of MPEG-7 characterizes color attributes of images [2]. A set of dominant colors in a region of interest or in an image provide a compact description that is easy to index. The target application is similarity retrieval in large image databases using dominant colors.

We propose to employ existing MPEG-7 dominant color descriptor of an image for region growing based segmentation. This opens another application for the dominant color descriptor. At the same time, it eliminates any additional computation required for adaptation of color similarity thresholds. Therefore, it speeds up automatic segmentation.

In the next section, the dominant color descriptor is explained. The centroid linkage method is summarized in the section 3. The section 4 presents a segmentation method that takes advantage of the dominant color descriptor.

2. DOMINANT COLOR DESCRIPTOR

The dominant color descriptor depicts part or all of an image using a small number of colors. For example, in a picture of a man dressed in a bluish shirt and reddish pants, blue and red would be the dominant colors of that person and the dominant color descriptor would contain not only those colors, but also the level of accuracy in depicting those colors within the given area. To compute this descriptor, the colors present in a given image or region are first clustered. This results in a small number of colors and the percentages of these colors are calculated. As an option, the variances of the colors assigned to a given dominant color are also computed. A spatial coherency value is also computed that differentiates between large color blobs versus colors that are spread all over the image. The difference between the dominant color descriptor and the color histogram descriptor is that the representative colors are computed from each image instead of being fixed in the color space, thus allowing the feature representation to be accurate as well as compact.

By successive divisions of color clusters with the generalized Lloyd algorithm [7] algorithm in between and then merging of the color clusters, the dominant colors are determined. This algorithm measures the distances of color vectors to the cluster centers, and groups the color vectors in the cluster that has the smallest distance. First, all color vectors $x(p)$'s are assumed to be in the same cluster C_1, i.e., the number of clusters is 1. For each cluster C_i, a color cluster center c_i is computed by means of averaging. After grouping the color vectors into the closest clusters, a distortion score D_i is computed for each cluster

$$D_i = \sum_p v(p) \|x(p) - c_i\|^2 \qquad x(p) \in C_i \tag{1}$$

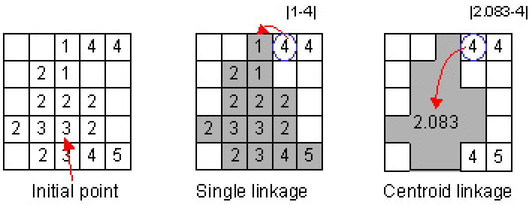

Fig. 1. Centroid-linkage algorithm compares a candidate pixel with the centroid. Note that single-linkage exclude the encircled pixel but it allows a pixel that has even higher value.

where c_i is the centroid of cluster C_i, and $v(p)$ is the perceptual weight for pixel p. The perceptual weights are calculated from the local pixel statistics to account for the fact that human vision perception is more sensitive to changes in smooth regions than in textured regions. The distortion score is the sum of the distances of the color vectors to the cluster center, and it measures the number of color vectors that changed their clusters at each iteration. The extraction problem is formulated as one of the minimizing the distortion in each cluster iteratively. The grouping is repeated until the distortion difference becomes negligible.

Then, each color cluster is divided into two new cluster centers by perturbing the center if the number of total clusters are less than a maximum, which is power of 2. As a final stage, the clusters that have close color centers are grouped to decide a final number of the dominant colors c_i's where $i = 1..N$. In Fig. 3, sample quantized images using the dominant colors are demonstrated. The color space is chosen as the RGB, however CIE-Lab and other color spaces are also considered in the standard. In [2], an average of 4-5 dominant colors were selected to represent regions.

3. CENTROID-LINKAGE GROWING

In principle, growing methods are applicable whenever a distance measure and linkage strategy can be defined. Several linkage methods were developed, they distinguished in the spatial relation of the pixels for which the distance measure will be computed.

Suppose that we start with a single pixel p and wish to expand from that seed pixel to fill a coherent region. Let's define a distance measure $\Psi(p, q)$ such that it produces a low value if pixels p and q are similar and a high value otherwise. Now, consider a pixel p is adjacent to pixel another pixel q. We can include pixel q into pixel ps region if distance $\Psi(p, q) < \epsilon$ for some threshold ϵ. We can then proceed to the other neighbors of p and do likewise. Suppose that $\Psi(p, q) < \epsilon$ and we added pixel q to pixel ps region. We can now similarly consider the neighbors of q and add them likewise if they are similar enough. Of course, we now have unanswered questions to address:

- How do we define distance measure Ψ?
- What threshold ϵ do we use?
- How do we update region attributes?

Several linkage methods were developed, they distinguished in the spatial relation of the points for which the distance measure will be computed. One obvious distance measure is to compare individual pixels color values. In centroid-linkage, a pixel p is compared to a region-wise centroid r_i by evaluating distance $\Psi(r_i, p)$ between the centroid of the target region and the pixel as in Fig.1. Centroid-linkage prevents from possible region leakages in case of image intensity is smoothly changing and strong edges that encircles regions are missing. It can construct a homogeneous region without detectable edge boundaries, although this property sometimes causes segmentation of a smooth region with respect to initial parameters. The norm of the distance measure should reflect significant intensity changes into the distance magnitude and suppress small variances.

The threshold of distance measure determines how homogeneous a region should be at the end. Small thresholds tend to generate multiple color consistent but smaller regions. Such thresholds causes over-segmentation. On the other hand, larger thresholds may combine different colored region into one by passing-over weak edges, which brings in under-segmentation. Generally, under-segmentation requires more effort to cope with than over-segmentation. In a way, distance threshold controls the colors variance of the region without explicitly computing it. Color dynamic range is another constrain effects threshold.

One simple centroid statistic is to keep an updated mean of the region pixels. As each new pixel is added the mean is updated. Although gradual drift is still possible, the weight of all previous pixels in the region act as a damper on such drift.

Initially, a region consists of a selected seed pixel alone. A similar technique is to initialize the region with not only a single pixel but a small set of pixels to better describe the regions statistics. With such initialization, not only a region mean is suggested but the variance as well. Candidate pixels can be compared to the region mean with respect to the region variance. The variance can be computed by sampling a small area around the initial seed pixel.

4. APPLICATION TO SEGMENTATION

Regions are grown from the seed pixels. A seed pixel should be selected such that it can characterize its local neighborhood as relevant as possible. Pixels have small color gradient magnitude are good candidates to represent their local neighborhood. Thus, the color gradient magnitude $|\nabla I(x, y)|$ is computed, and the minimum gradient magnitude pixels are chosen as seeds s_i iteratively. For computational simplicity, gradient may be computed only for a down-sampled image. A minimum gradient pixel is selected as seed pixel s_i, a region R_i is initiated, and region's centroid r^i is set to the seed pixel's color values. Then the adjoint pixels p are evaluated in 4-pixels neighborhood. Color distance $\Psi(s_i, p)$'s are computed. If the color distance is less than a threshold $\Psi(s_i, p) < \epsilon$, the pixel p is included in the region, assigned as an active pixel, and the centroid colors is updated by the averaged means.

After a region R_i is grown, all the pixels of the region R_i is removed from the set Q. The seed selection can be formulated as

$$s_i = \arg\min_{Q} |\nabla I(x, y)| \quad ; \quad Q = S - \bigcup_{j=1}^{i} R_j \qquad (2)$$

where Q is the set of all possible image pixels initially. The next minimum in the remaining set is chosen, and selection process is iterated until no more pixel remains.

Using the dominant colors, inter-color distances are computed for separate channels. The projections of dominant colors are ordered and represented as $l_k(i)$ where $k = r, g, b$ is the channel. Fig.2 illustrates the notation. For each channel, the dominant colors that have close projections are combined together. The definition of closeness is kept very strict, i.e. around $1/64$ of the dynamic color range, thus it does not affect the segmentation performance or causes over-segmentation. It is only used for the elimination of very similar projections in a channel. The number of projections in each channel i_r, i_g, i_b are obtained as a result. Note that there are multiple projections for a channel such that $l_k(i) < l_k(i+1)$ and $i = 1..i_k$ for color channel k where i_k is the number of projections.

For each projection $l_k(i)$, two values $l_k^-(i)$ and $l_k^+(i)$ that correspond to the distances from the nearest projections $l_k(i+1)$ and $l_k(i-1)$ on the both sides are determined

$$\begin{aligned} l_k^-(i) &= \frac{1}{2}(l_k(i) - l_k(i-1)), \\ l_k^+(i) &= \frac{1}{2}(l_k(i+1) - l_k(i)) \end{aligned} \tag{3}$$

Using the seed pixel, three lengths l_r, l_g, l_b are computed

$$l_k = \begin{cases} l_k^-(i) & l_k^-(i) < I_k(s) \leq l_k(i) \\ l_k^+(i) & l_k(i) < I_k(s) \leq l_k^+(i) \end{cases} \tag{4}$$

These lengths represent the inter-cluster distances between the dominant colors at the corresponding color channel, and are utilized to determine the Lorentzian-based distance measure between the centroid and the candidate pixel p

$$\Psi(r,p) = \sum_k i_k \log(1 + \frac{|I_k(r) - I_k(p)|}{l_k}) \tag{5}$$

where $k = r, g, b$. The Lorentzian term is chosen since it is sensitive enough towards the small color differences while it prevents the computed distance from bursting for relatively large color difference in a single channel. Similar to the analogy for a robust estimator, it does not amplify color distance linearly or quadratically. In contrast, when the magnitude of the of the distance is small the function ψ increases moderately but then it keeps same for extremely deviant distances.

The total number of dominant colors in a channel is multiplied with the distance term to increase the contribution of a channel that supplies more details, i.e. multiple dominant colors for segmentation. The distance threshold is given as

$$\epsilon = (i_r + i_g + i_b) \tag{6}$$

which means that the $I_r(p)$, $I_g(p)$, $I_b(p)$ are all within the same color bins with the centroid.

It should be reminded that the above computation is insignificant since it involves only very small number of dominant colors.

Since no prefiltering of the input image is utilized to reduce the computational complexity, some of the regions may be negligible in size due to the noise and edges. Therefore, the small regions are removed as a final step.

5. DISCUSSION

We simulated the algorithm using several images from the MPEG-7 test sequences. In order to make a fair comparison between the proposed method and other state-of-the-art region growing methods, we designed two other algorithms. The first algorithm utilizes an L^2 color distance norm as commonly employed in several previous work. The distance threshold is set to an experimentally derived value ($3 * 2^{-10}$ after color range normalized to [0,1]) for all input sequences. This value gives the best overall segmentation performance for the several images that have been exhaustively tested. The second algorithm uses an adaptive threshold for each separate image unlike the first one. This threshold is obtained by the variances of the histograms and tuned to obtain best possible segmentation results. The distance metric is chosen as the magnitude norm instead of the L^2. Both of the test algorithms uses region growing method with the initial seeds points chosen as the proposed method. To favor the comparison algorithms, the input images are median filtered to remove noise for these algorithms, whereas not filtered for the proposed method. Only very small regions are removed from the results of the proposed methods. We summarized some of the results obtained with all three algorithms in Fig.4. For clarification, each segmented image is coded with random colors.

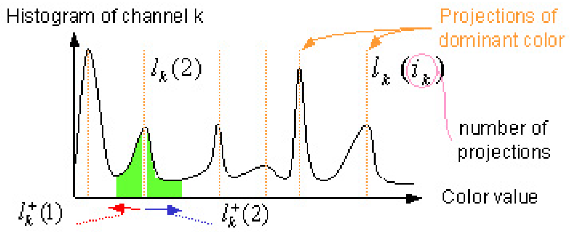

Fig. 2. The histograms are used to determine color modalities of video to decide thresholds.

As visible in the given test results, the adaptation method performs significantly better. It determines the distance threshold for every image without any complex computation, and does not cause over or under segmentation even the input image is not filtered. Initial filtering is a prerequisite for most algorithms especially the thresholds are adapted locally for every pixel.

The region growing method is obviously faster than the watershed methods and morphological operators.

Using the MPEG-7 dominant color descriptor to automatically adapt segmentation threshold is a novel method.

6. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 N4031, Coding of Moving Pictures and Audio, Singapore, March 2001

[2] B.S. Manjunath, J.R. Ohm, V. Vasudevan, and A. Yamada, "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 6, June 2001

[3] R. Taylor and P. Lewis, "Color Image Segmentation Using Boundary Relaxation", ICPR, Vol.3, 721-724, 1992

[4] F. Meyer, "Color image segmentation", ICIP, Netherlands, 303-304, 1992

[5] L. Priese and V. Rehrmann, "A fast hybrid color segmentation method", DAGM, 297-304, 1993

[6] T. Westman and D.Harwood, "Color Segmentation by Hierarchical Connected Components Analysis with Image Enhancements", ICPR, Vol.1, 796-802, 1990

[7] M. Sabin, "Global convergence and empirical consistency of the generalized Lloyd algorithm", PhD thesis, Stanford University, 1984

Fig. 3. The first column is the original images, and the other columns are the quantized images using the dominant colors. The number of dominant colors is set to 32, 16, 8, 4 respectively. The dominant colors are shown next to each image. Using dominant colors the range of the colors can be effectively reduced from 2^{24} for an 8-bits coded color image to a small number without corrupting image. The quantized images by the dominant colors can be employed an initial stage of image segmentation, although in this contribution we investigated its effect on region growing.

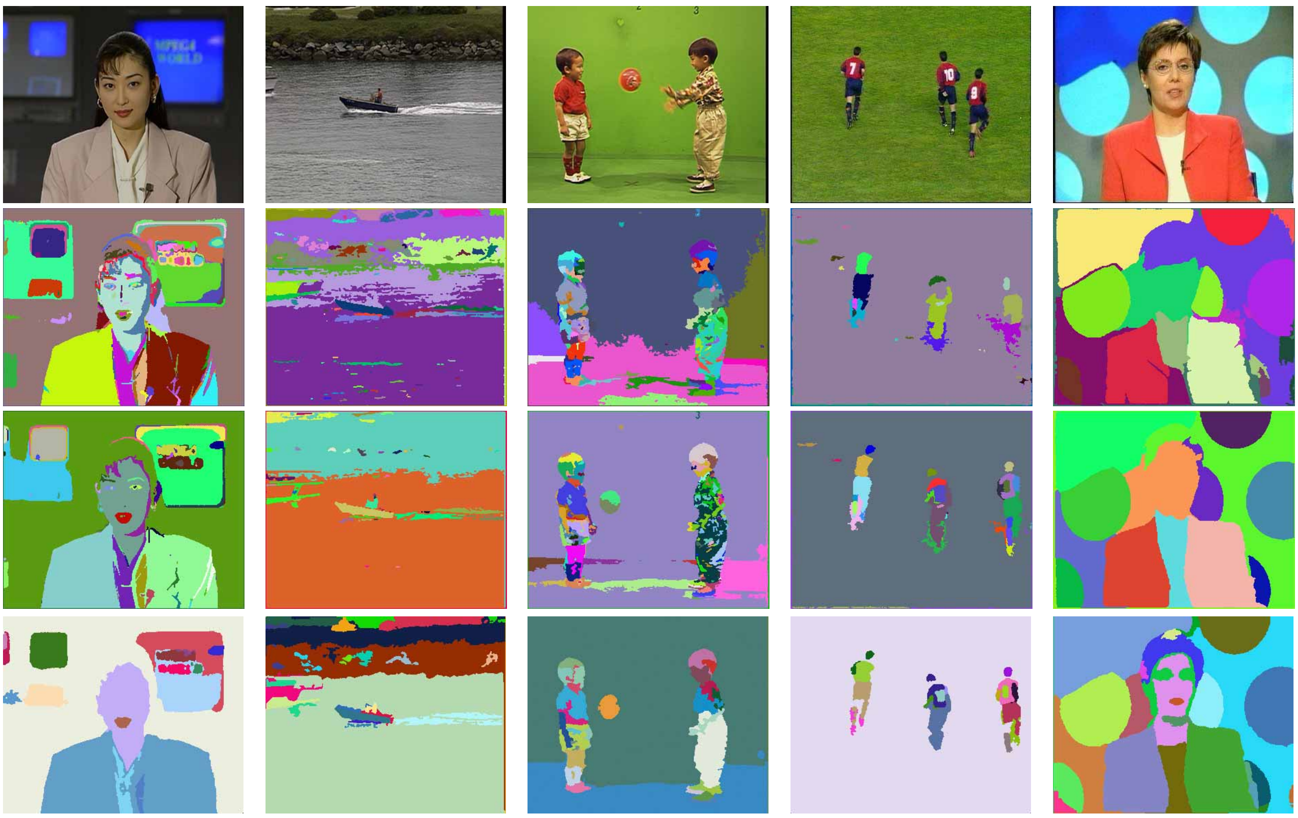

Fig. 4. The first row shows the original test images. The second row is the result of the region growing with constant threshold and magnitude-square distance measure. The third row presents the segmented images by variance based adaptive thresholds and magnitude distance measure. The last row shows the results obtained by the proposed algorithm. Note that, the $Akiyo$ (cloth), $Child$ (background) are over-segmented whereas $Futbol$ (head-arm), $Speaker$ (face) are under-segmented in the both comparison methods. The proposed threshold adaptation method gives better performance even no parameter tuning was done.