VIDEO SEGMENTATION USING VECTOR-VALUED DIFFUSION AND CLUSTERING

Daehee Kim[1], *Chung-Hyun Ahn*[1] *and Yo-Sung Ho*[2]

[1]Electronics and Telecommunications Research Institute (ETRI)
161 Gajeong-dong Yuseong-gu, Daejeon, 305-350, Korea
{daeheekim, hyun}@etri.re.kr
[2]Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-dong Buk-gu, Kwangju, 500-712, Korea
hoyo@kjist.ac.kr

ABSTRACT

In this paper, we propose a user-assisted video segmentation algorithm based on color information to alleviate over-segmentation problems. We perform intra-frame segmentation by image simplification, region labeling, and color clustering. In this paper, we also present a discrete three dimensional diffusion model for easy implementation. The statistical property of each labeled region is used to estimate the number of total clusters, and agglomerative hierarchical clustering is performed with the estimated number of clusters. Since the proposed clustering algorithm counts each region as a unit, it does not generate oversegmentation problems along region boundaries. For inter-frame segmentation, we employ a look-up table for foreground color clusters, track the foreground regions, and utilize those information to extract moving objects.

1. INTRODUCTION

The MPEG-4 visual coding standard [1] enables content-based functionalities by introducing the concept of the video object plane (VOP). However, since it is not easy to define a mathematical model and a similarity measure for extracting video objects adequately, automatic segmentation algorithms cannot provide satisfactory segmentation results over various image sequences. Therefore, the user-assisted segmentation approach is more practical in generating VOPs of moving objects. In this paper, we propose a new image segmentation tool based on region information, which allows the user to collect the segmented regions for constructing meaningful objects.

In the proposed algorithm, the intra-frame segmentation consists of three operations: image simplification, region labeling, and color clustering, while the inter-frame segmentation exploits foreground and background color information. According to Skarbek's classification [2], our proposed algorithm can be classified as a mixture of pixel-based segmentation and region-based segmentation. In this paper, we propose a discrete 3-D diffusion model with a shock filter and a new clustering method that does not require any priori known information. In order to extract moving objects in the inter-frame segmentation step, we make a loop-up table for foreground colors, track foreground regions, and utilize those information.

2. VECTOR-VALUED DIFFUSION PROCESS

In order to simplify region textures, we should find an image smoothing operation in homogeneous regions, but not across region boundaries. This requirement can be satisfied with an anisotropic diffusion filter [3]. The diffusion is carried out in the perceptually uniform LUV color space. If each component in the LUV color space is diffused independently, it evolves with different smoothing directions and intensities. Therefore, the vector-valued diffusion is performed in the 3-D LUV color space by

$$\frac{\partial \mathbf{I}(x,y,t)}{\partial t} = div(\rho(x,y,t)\nabla \mathbf{I}(x,y,t)) \qquad (1)$$

where $\mathbf{I}(x,y,t)$ is a vector in the LUV color space and ρ is a diffusion or conduction coefficient.

The scalar-valued diffusion and the vector-valued diffusion require normal and tangential directions of gradients [4]. In order to alleviate these problems, Perona-Malik's resistor model [3] can be employed; however, this model was originally developed for scalar-valued images. In this model, each pixel is connected to four neighboring pixels with resistors whose resistances increase as the pixel differences of each connection increase.

In this paper, we extend Perona-Malik's resistor model to vector-valued images. Since the diffusion equation relates the amount of the temporal variation to the amount

This work was supported in part by K-JIST, in part by KOSEF through UFON, and in part by MOE through BK21.

of the spatial variation, it is more reasonable to employ the eight-connected resistor model rather than the four-connected resistor model. In addition, we describe resistances between each connection with the local variation $|d\mathbf{I}|$ for vector-valued images.

Therefore, the discretized 3-D diffusion equation of Eq. (1) is described by

$$\mathbf{I}_{y,x}^{t+1} = \mathbf{I}_{y,x}^{t} + \frac{1}{8}\sum_{i=1}^{4} r_i d\mathbf{I}_i + \frac{1}{8\sqrt{2}}\sum_{i=5}^{8} r_i d\mathbf{I}_i \qquad (2)$$

where r_X is an admittance function of each direction, and $d\mathbf{I}_X$ represents eight neighbor differences. To compensate for different distances from pixels to their neighbors, we scale the differences of the diagonal directions by $1/\sqrt{2}$.

3. REGION LABELING CLUSTERING

Some color segmentation algorithms have performed clustering in a color space directly; however, those algorithms show coarse segmentation results near object boundaries, because those algorithms use only the color histogram and the similarity measure; it does not consider the region information of objects properly.

In order to alleviate these problems, we employ a gradient-based watershed algorithm and obtain partitioned regions of the simplified image that was obtained from Section 2. In this work, we use an immersion-based watershed algorithm [5]. The color gradient image, the input image to the watershed algorithm, is generated by $||d\mathbf{I}||^2$.

Our goal here is to find the best representative clusters in the LUV color space and to preserve the boundaries of objects. In the proposed method, we estimate the number of color clusters based on the region information. The proposed clustering algorithm consists of two steps: estimation of the number of clusters, and agglomerative hierarchical clustering.

In order to estimate the number of clusters required, we assume that each vector component in each region resulted from the watershed algorithm obeys a Gaussian distribution $N(\mu_i, \sigma_i)$, where i indicates one component among the LUV color space. This assumption is valid because a vector in each diffused region has a small Euclidian distance to the mean vector in the corresponding region. Under the hypothesis H_0 that two points are in the same region, the difference d_i between two points in the same region obeys a zero mean Gaussian distribution $N(0, \sqrt{2}\sigma_i)$.

Since the Euclidian distance is a good representation of the color distance in the LUV color space that is perceptually uniformly distributed, we consider the normalized test statistic

$$\theta = \sum_{i=1}^{3} \frac{d_i^2}{2\sigma_i^2} \qquad (3)$$

Algorithm 1: Estimation of Number of Clusters

begin initialize n, $\mathbf{T}_1 \leftarrow R_i$, $R_i \leftarrow \mathbf{m}_i$, $c \leftarrow 1$, $i \leftarrow 1$
  do $i \leftarrow i+1$
    find nearest T_j to R_i among c reference regions
      based on Euclidian Distance
    compute θ
    if$(\theta > \theta_{th})$
      $c \leftarrow c+1$
      $\mathbf{T}_c \leftarrow R_i$
    else
      merge $\mathbf{T}_j \leftarrow R_i$ and $\mathbf{T}_j$
      recompute $\mathbf{m}_j$ on $\mathbf{T}_j$
      recompute variances of each space on $\mathbf{T}_j$
  until $i = c$
  return c
end

Fig. 1. Estimation of number of clusters

where i is the index for each space of the LUV color space. In Eq. (3), the test statistic θ has a χ^2 probability density function with three degrees of freedom.

$$p(\theta|H_0) = \frac{1}{\sqrt{2\pi}}\theta^{\frac{1}{2}} exp\left(-\frac{\theta}{2}\right) \qquad (4)$$

With the known distribution $p(\theta|H_0)$, the decision on whether or not two points are in the same region can be made by a significance test [6]. For this purpose, we specify a false alarm rate α.

If we obtain n partitioned regions from the watershed algorithm, we assign a central vector to each region. A central vector is a representative vector of each region, which is the average vector of the corresponding region. The test statistic θ is now evaluated by the differences between the central vector of a certain region to be tested and the central vectors of the reference regions where the test statistic θ is normalized by variances of the reference regions. Whenever it exceeds θ_{th}, this region is declared to be distinct from the previous reference regions. Thus, the estimated number of clusters is increased by one.

Fig. 1 shows the estimation procedure to find the number of clusters. The reference regions are updated and the final number of the reference regions is the estimated number of clusters. All regions obtained by the watershed algorithm are sorted by the region size in the descending order. The largest region is the initial reference region $\mathbf{T}_1$ and each region R_i is represented with the mean vector $\mathbf{m}_i$. In Fig. 1, each region $\mathbf{T}_j$ is the reference region, c is the estimated number of clusters and n is the number of regions obtained from the watershed algorithm.

After the number of clusters is estimated, we merge the regions generated from the watershed scheme by an agglomerative hierarchical clustering algorithm [7]. There-

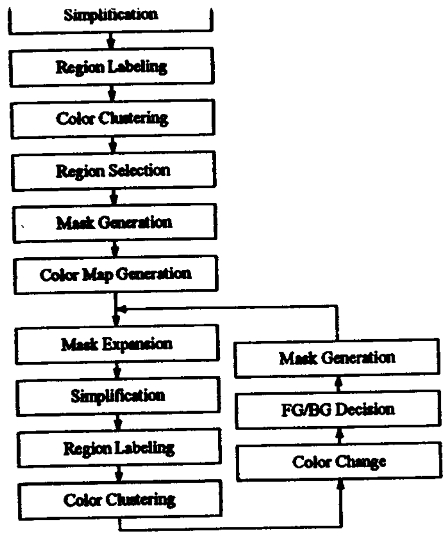

Fig. 2. User-assisted segmentation algorithm

fore, the final color clusters segment images based on colors and regions; however, the proposed clustering algorithm does not require any a priori information. Since this clustering algorithm reflects the shape of the region, it does not generate oversegmentation along region boundaries.

4. SEGMENTATION FOR VIDEO SEQUENCES

In this section, we explain a video segmentation procedure. As shown in Fig. 2, the intra-frame segmentation consists of image simplification, region labeling, color clustering, region selection, mask generation, and color map generation. In the proposed algorithm, each region can be represented by a single color; therefore, each region is so large that the user can handle it easily.

The region selection step requires user's interaction to obtain user's cognitive information. In this step, we select the regions using the user-pointing devices. These selected regions comprise an interesting object. After an object is defined by user's selection, we generate the mask from the selected regions.

Due to the color clustering step, an image is represented by several representative colors. In the color map generation step, we make two look-up tables from the representative colors. One of them is comprised of colors inside selected regions. The other look-up table consists of the rest representative colors.

The inter-frame segmentation operation comprises mask expansion, image simplification, region labeling, color clustering, color change, FG/BG decision, and mask generation. The inter-frame segmentation is performed repeatedly until the end of the image sequence. Here, the simplification, the procedures as in the intra-frame segmentation step.

In the mask expansion step, we check the motion activity of foreground regions and find the maximum offset between the current and previous frames. After we find the maximum offset, we expand the previous segmented mask where the amount of the expansion is proportional to the absolute value of the maximum offset. The expansion of the mask, which is a binary image, is performed by the morphological dilation operation.

Here, we assume that an object consists of several colors and these colors are not changed until the end of the image sequence. In the color change step, representative colors of the current frame are replaced by those of the previous frame. A representative color of the current frame is compared with two look-up tables. We find the nearest color among two look-up tables using the Euclidian distance. Therefore, all regions of the current frame are represented by the look-up tables of the previous frame.

In the FG/BG decision step, we consider all regions inside the expanded mask as the candidates of the foreground regions. Among those candidate regions, we select foreground regions if their colors are the same as the colors of the look-up table for foreground regions in the previous frame. Although regions outside the expanded mask have the foreground colors in the previous frame, those regions are considered as background. In the mask generation step, we generate the final segmentation mask by collecting foreground regions and then we update the foreground color look-up table and the background color look-up table. Since the intensity of colors may change through time evolution, this update is necessary despite of the assumption that object colors are not changed. In addition, since this method considers colors and region boundaries, it can accommodate topological changes.

5. EXPERIMENTAL RESULTS

In order to evaluate the proposed user-assisted segmentation algorithm using color clustering, we have performed computer simulations on various kinds of test images. Fig. 3 shows the simulation results from the MOTHER AND DAUGHTER image. After we applied morphological filtering, anisotropic diffusion, watershed, estimation of number of color clusters, and agglomerative hierarchical clustering operations on the sequence, we obtained results in Fig. 3.

While Fig. 3(b) is the diffused image by Eq. (2) without the morphological filtering, Fig. 3(c) is the diffused image after morphological filtering. In Fig. 3(c), we can ignore specific detail image structures, which are not interesting for semantic image analysis because of small regions. Fig. 3(d) and Fig. 3(e) are output images by the watershed algorithm on the diffused image without morphological fil-

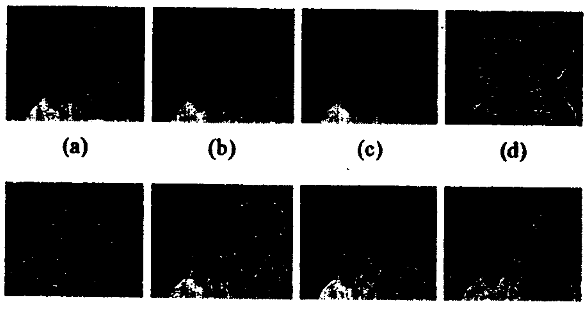

Fig. 3. Results for MOTHER AND DAUGHTER

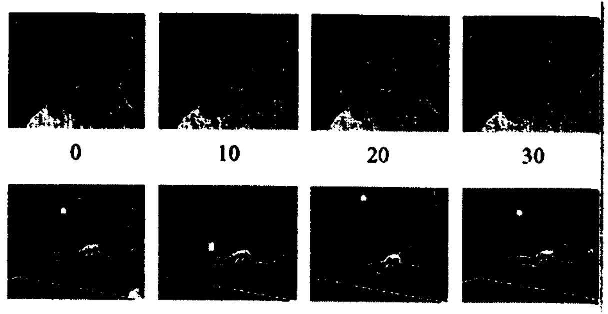

Fig. 4. Video segmentation

tering and with morphological filtering, respectively. We note that Fig. 3(e) is much simpler than Fig. 3(d).

Fig. 3(f) shows the segmentation result based on the clustering using a color histogram [8]. Fig. 3(f) demonstrates oversegmentation at region boundaries because the clustering algorithm has directly been applied on the diffused or low-pass filtered images in a certain color space and the diffused or low-pass filtered images can generate the mixed colors along the region boundaries.

On the other hand, the final result of our proposed scheme does not generate oversegmentation at region boundaries and it can ignore the unwanted details, as shown in Fig. 3(g). Algorithm 1 shown Fig. 1 estimates that the MOTHER AND DAUGHTER image has twelve color clusters and the clustering algorithm makes twelve color clusters. Fig. 3(h) displays the segmented image that is represented by the final twelve color clusters.

Fig. 4 shows results of video segmentation using color information for MOTHER AND DAUGHTER and TABLE TENNIS sequences. The first pictures of each row in Fig. 4 are results of the intra-frame segmentation. The first picture of in the first row is generated by merging selected regions from among regions of Fig. 3(g), and the first picture of the first row results from TABLE TENNIS sequence using the same precedure. The rest pictures are results of the inter-frame segmentation.

Although the proposed inter-frame segmentation is quite simple, it works well on image sequences because most images consist of several dominant colors. In addition, since the proposed algorithms consider only regions with foreground colors, it can accommodate topological changes.

6. CONCLUSIONS

In this paper, we have presented a new region-based color segmentation algorithm. In order to take advantage of the perceptually uniformly distributed LUV color space, we perform the vector-valued diffusion process in the LUV color space. In addition, statistical characteristics of each labeled region are employed to estimate the number of total clusters, and agglomerative hierarchical clustering is performed with the estimated number of clusters. Therefore, the proposed clustering algorithm does not require any a priori information. Since the proposed clustering algorithm reflects the shape of the region, it does not generate oversegmentation problems along region boundaries. In addition, we employ a loop-up table for foreground colors, track the foreground regions, and utilize those information to extract moving objects in the inter-frame segmentation step. Although the proposed inter-frame segmentation is quite simple, it works well on various image sequences because most images consist of several dominant colors.

7. REFERENCES

[1] ISO/IEC FDIS 14496-2: "Information technology - generic coding of audio-visual objects, Part 2: visual," *ISO/IEC JTC1/SC29/WG11*, Oct. 1998.

[2] W. Skarbek and A. Koschan, "Colour image segmentation - A survey," *Technical Report 94-32, Technical University of Berlin*, Oct. 1994.

[3] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. PAMI*, vol. 12 no. 7, pp. 629-638, July 1990.

[4] G. Sapiro and D. L. Ringach, "Anisotropic Diffusion of Multivalued Images with Applications to Color Filtering," *IEEE Trans. Image Processing*, vol. 5, no 11, pp. 1582-1585, Nov. 1996.

[5] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. PAMI*, vol. 13, no. 5, pp. 583-598, June 1991.

[6] D.C. Montgomery, *Design and Analysis of Experiments 5th Edition*, John Willy & Sons Inc., New York, 2001.

[7] R.O. Duta, P.E. Hart and D.G. Stork, *Pattern Classification*, John Wiley & Sons Inc., Singapore, 2001.

[8] L. Lucchese and S.K. Mitra, "Color image segmentation through independent anisotropic diffusion of complex chromaticity and lightness," *Proc. of ICIP'2001*, vol. 1, pp. 746 -749, 2001.