Semi-Automatic Segmentation for Video Coding

Minho Kim and Yo-Sung Ho
Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-Dong Puk-Gu, Kwangju, Korea
E-mail: mhkim@gogh.kjist.ac.kr, hoyo@kjist.ac.kr

Abstract

This paper describes a new semi-automatic segmentation algorithm based on color information. Semi-automatic segmentation mainly consists of intra-frame segmentation and inter-frame segmentation. While intra-frame segmentation extracts video objects of interest from boundary information provided by the user and intensity information of the image, inter-frame segmentation partitions the image into the video objects and the background by tracking the motion of video objects. For inter-frame segmentation, color information (Y, Cb and Cr) can be used efficiently to find the accurate boundary of video objects. In this paper, we propose a new region growing algorithm that maximizes the ability of region differentiation.

1. Introduction

Contrary to the conventional block-based image coding standards, such as MPEG-1 and MPEG-2, the new MPEG-4 standard is targeting for content-based coding where each frame of the video sequence is represented by video object planes (VOPs). Therefore, object segmentation is required to obtain VOPs from each picture frame in MPEG-4.

For object-based image coding, several algorithms for automatic segmentation have been developed: change detection-based segmentation [1], motion-based segmentation [2], morphological segmentation [3,4], and spatio-temporal segmentation [5].

Change detection-based segmentation employs a global thresholding operation. A critical factor in change detection is the decision threshold. Aach and Kaup determined the decision threshold empirically by relating to the false alarm rate [1]. However, this method fails to distinguish the covered and uncovered background from video objects.

Motion-based segmentation [2] solved the covered and uncovered background problem by assuming that the area to be segmented is defined by a set of uniform motion and position parameters. These mapping parameters can be estimated from spatio-temporal differences of pixel intensity values. However, this approach still contains some problems due to the global thresholding operation.

While the above two approaches utilize temporal information by taking differences between two successive picture frames, morphological segmentation employs mathematical morphology operations which are very useful for segmentation [3,4]. Mathematical morphology can efficiently deal with geometric features, such as size, shape, contrast and connectivity, which can be considered as object-oriented features. Thus, the morphological segmentation algorithm can preserve boundaries of video objects. However, each partitioned region itself may not represent a meaningful video object.

Recently, there have been several attempts to solve the segmentation problems mentioned above by exploiting both temporal and spatial information [5]. However, they could not solve the problems completely, because the definition of the video object is very subjective and automatic segmentation methods use only intensity information contained in the video frames. This implies that video segmentation needs some kind of user interaction for the purpose of improving the performance of segmentation results. This is the main motivation for semi-automatic segmentation [6].

In this paper, we propose a new algorithm for semi-automatic segmentation based on color information. For the first picture frame, we extract video objects of interest with the boundary information provided by the user. We use a double labeling method to utilize the user information efficiently [7]. For subsequent picture frames, we extract video objects by tracking the motion of video objects [6].

During video object tracking, we may make some mistakes. In order to compensate for these errors, we apply a boundary fitting algorithm based on region growing. Conventional algorithms for semi-automatic segmentation show some mistakes due to feature blurring of color components.

In this paper, we also propose a new region growing algorithm that exploits color information efficiently for semi-automatic segmentation. By preventing the boundary between two distinct regions from being merged, we can obtain better segmentation results.

2. Semi-Automatic Segmentation

Semi-automatic segmentation consists of intra-frame segmentation and inter-frame segmentation. In the first picture frame, we extract video objects of interest with morphological segmentation tools. Then, we apply inter-frame segmentation for the subsequent picture frames until we encounter one of the following three conditions: (a) appearance of new video objects, (b) scene changes, or (c) unsatisfactory segmentation results. If one of the three events occurs, we restart the intra-frame segmentation operation.

2.1. Intra-Frame Segmentation

By intra-frame segmentation, we partition the image into semantically meaningful regions. In other words, the image is divided into the background and video objects. Intra-frame segmentation is composed of global labeling, local labeling, and merging the global and the local labeling.

Global Labeling by User Interaction

For the global labeling, the user can provide the boundary information of video objects by indicating video objects of interest. In other words, the user can draw the uncertainty region around boundaries of video objects manually. This globally labeled mask contains the background, video objects and uncertainty areas.

Local Labeling by Morphological Segmentation

We generate a locally labeled mask from the intensity information of the image itself. Since the morphological segmentation algorithm exploits the topographic characteristics of the image, it preserves the boundary of the video object well. However, other region labeling algorithms need some adjustment of segmentation parameters, such as k values in the k-means algorithm, to produce reasonable segmentation results for various kinds of video sequences.

In this paper, we adopt the morphological segmentation algorithm for local labeling. The morphological segmentation algorithm divides the image into small regions of similar intensity values. The segmentation result looks like a mosaic. Thus, we call the result produced by morphological segmentation as a locally labeled mask.

The morphological segmentation algorithm [3] contains three main operations: (a) image simplification, (b) marker extraction, and (c) region decision.

Image simplification is accomplished by morphological filters, especially by morphological opening/closing by reconstruction. We utilize these reconstruction filters, instead of simple morphological opening and closing, because reconstruction filters allow a good preservation of contour information while simple opening and closing cause a loss of contour information. The morphological opening by reconstruction removes bright details which are not fitted within the structuring element, and the morphological closing by reconstruction removes dark details.

Marker extraction is achieved by the morphological gradient operator, which is defined by

$$g = \delta(f) - \varepsilon(f) \tag{1}$$

where $\delta(f)$ is a dilation operator, and $\varepsilon(f)$ is an erosion operator. The marker area corresponds to the region whose pixels have low morphological gradient values.

We then carry out the region decision operation using the watershed algorithm [4]. The watershed algorithm is originated from the topography, which deals with catchment basins and their dividing lines, called as watershed lines. Watershed lines partition the image by associating an area (catchment basin) to each local minimum, which corresponds to the locally lowest morphological gradient value.

Merging of Globally and Locally Labeled Masks

The globally labeled mask is composed of video objects, the background and uncertainty areas. This mask contains the uncertainty areas, but each partitioned region and video object have the property of one-to-one correspondence.

On the contrary, the locally labeled mask does not have that property. However, in each region, boundaries of video objects are well preserved because the image is partitioned by the topographic property of the morphological gradient operator.

By combining the two masks, we take advantage of the one-to-one correspondence property of the globally labeled mask and well-preserved boundaries of the locally labeled mask [7].

2.2. Inter-Frame Segmentation

Inter-frame segmentation extracts video objects from the background by tracking the video objects from the previous frame, as explained in Fig. 1. Inter-frame segmentation consists of object prediction, followed by boundary fitting.

For object prediction, we project each video object of the previous frame into the current frame by motion information of the video object. The motion of each video object is defined by the following six-parameter affine model.

$$\begin{aligned} x' &= a_1 x + a_2 y + a_3 \\ y' &= a_4 x + a_5 y + a_6 \end{aligned} \tag{2}$$

where the parameters, from a_1 to a_6, are estimated in the dense flow field [6].

Since object prediction often generates error regions due to inaccurate motion estimation and object occlusions, we should perform boundary fitting. For boundary fitting, we identify uncertainty areas along boundaries between two adjacent video objects. Each uncertainty area corresponds to the region of inaccurate motion estimation. In addition, we

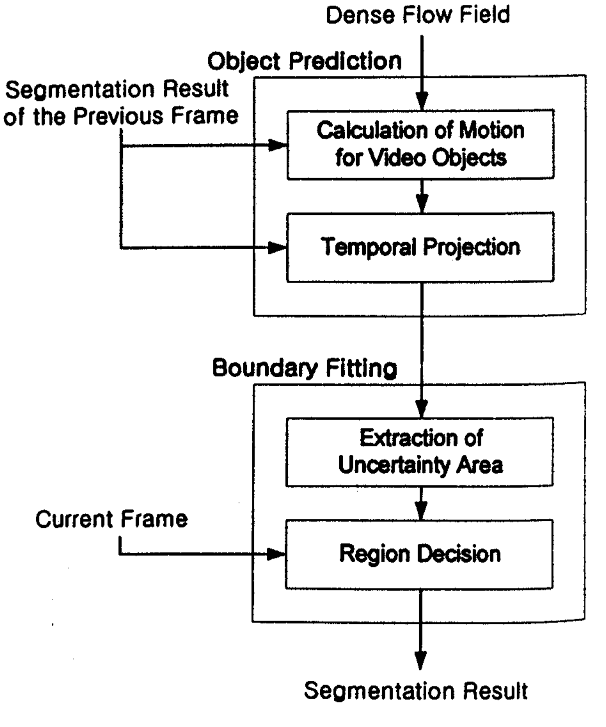


Fig. 1 Inter-Frame Segmentation

mark the unlabeled (uncovered) or multi-labeled (overlapped) pixels that are produced by temporal projection. We then seek the actual object boundary by the region growing algorithm that employs hierarchical queues.

Region Growing with Hierarchical Queues

The region growing algorithm for boundary fitting uses hierarchical queues to assign pixels of the uncertainty area to a corresponding video object. Hierarchical queues, as shown in Fig. 2, enable the region growing algorithm to decide the exact region for the uncertainty area by exploiting both proximity of pixel positions and similar characteristics of neighboring pixels simultaneously.

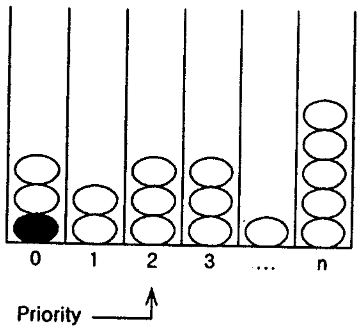


Fig. 2 Hierarchical Queues

The region growing algorithm consists of initialization and flooding. As shown in Fig. 3, the initialization operation inserts pixels of the uncertainty area, which are neighboring with video objects, into one of the hierarchical queues according to their priorities.

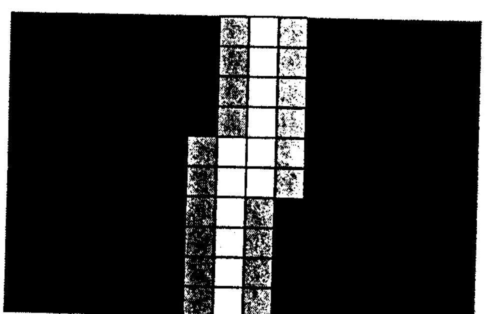


Fig. 3 Initialization for Region Growing

The flooding operation extracts a pixel in the queue with the highest priority (e.g., the gray pixel in Fig. 2) among the hierarchical queues that have at least one pixel. We assign the selected pixel to a video object that has the smallest *AbsDiff* value, defined in Eq. (4). Then, we insert uncertainty pixels that are neighboring with the selected pixel, if they are not yet inserted into the hierarchical queues. Fig. 4 illustrates this operation.

As explained in the flooding operation, pixels in the queue of the highest priority are treated first for region decision. Priority is defined by the following equations.

$$Priority = \min\{AbsDiff(n)\} \tag{3}$$

$$AbsDiff(n) = |CurrAvg - LAvg(n)| \tag{4}$$

$CurrAvg = \alpha * Crnt\_Y + \beta * Crnt\_Cb + \gamma * Crnt\_Cr$

$LAvg(n) = \alpha * LAvg\_Y(n) + \beta * LAvg\_Cb(n) + \gamma * Lavg\_Cr(n)$

α, β, γ: *weighting factors*, $\alpha+\beta+\gamma=1$

LAvg_Y, Cb, Cr: local averages in observation window centered on the current pixel

Crnt_Y, Cb, Cr: values of the current pixel

n: identifier for a video object

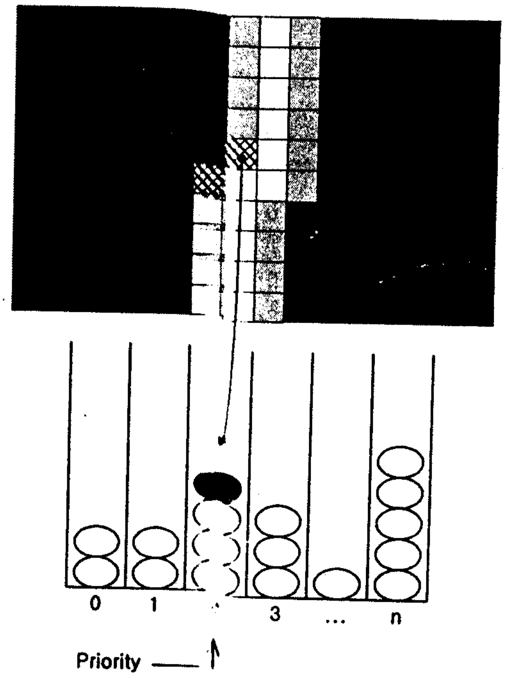


Fig. 4 Flooding in Region Growing

The above method, however, does not preserve individual features of each color component because of the averaging operation, and may produce inaccurate boundaries between two adjacent video objects even where their color difference is definite. This drawback can be observed in segmentation results.

In Eq. (5), we define a new difference measure that can increase the ability of region differentiation, while preserving the color information.

$$AbsDiff(n) = \max(|Diff\_Y(n)|, |Diff\_Cb(n)|, |Diff\_Cr(n)|) \tag{5}$$

$$Diff\_Y(n) = Crnt\_Y - LAvg\_Y(n)$$
$$Diff\_Cb(n) = Crnt\_Cb - LAvg\_Cb(n)$$
$$Diff\_Cr(n) = Crnt\_Cr - LAvg\_Cr(n)$$

In Eq. (5), we take the maximum value among all differences of color components. We make use of *AbsDiff(n)* in Eq. (5) not only as priority when we insert pixels in the uncertainty area, but also as a similarity measure when we assign the pixel extracted from the hierarchical queues into a region. Thus, we can locate an accurate boundary between video objects even where the difference of the luminance components (Y component) is small and the differences of color components (Cb or Cr) are large.

3. Experimental Results

In order to evaluate the performance of the proposed segmentation algorithm with a new difference measure, we select "Mother & Daughter" sequence, which has the QCIF format of 176x144 pixels (Fig. 5).

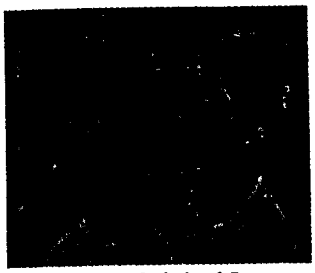
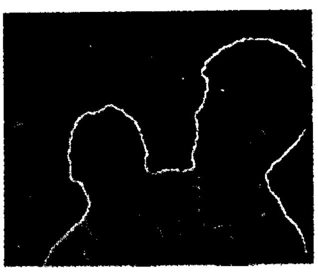


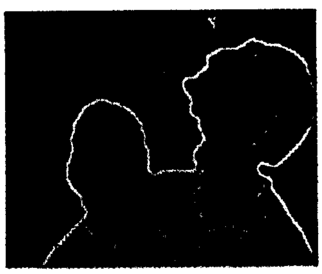
Fig.5 The Original Image

Fig. 6 shows segmentation results obtained by changing α, β, and γ values in order to assess the performance of the conventional algorithm (Eq. (4)). When we use only the Y component, the boundary of the left face and hair of Daughter collapsed (Fig. 6 (a)). On the contrary, when we exploit only the Cb or Cr component, the boundary of Mother collapsed, but the boundary of Daughter was preserved well (Fig. 6 (b) and (c)).

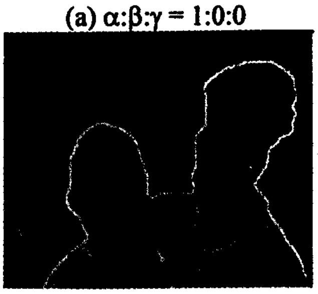
As we increase α from 0 to 0.33 (Fig. 6 (d), (e), and (f)) and decrease β and γ, the boundary of Mother is identified more accurately, while the boundary of Daughter collapses more severely.



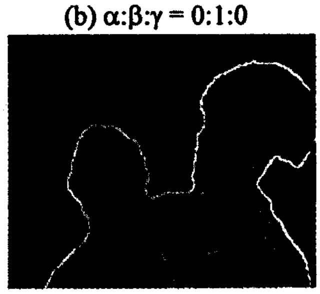
(a) $\alpha:\beta:\gamma = 1:0:0$



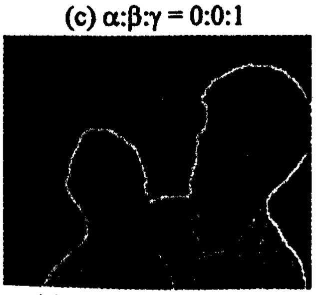
(b) $\alpha:\beta:\gamma = 0:1:0$



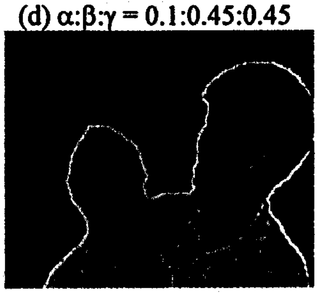
(c) $\alpha:\beta:\gamma = 0:0:1$



(d) $\alpha:\beta:\gamma = 0.1:0.45:0.45$



(e) $\alpha:\beta:\gamma = 0.2:0.4:0.4$



(f) $\alpha:\beta:\gamma = 0.33:0.33:0.33$

Fig. 6 Results by the Conventional Method

Fig. 7 Result by the Proposed Method

Segmentation results in Fig. 6 imply that individual colo component is not sufficient for locating the object boundar accurately and that the averaging operation is not adequat for mixing color information.

Fig. 7 shows a segmentation result by our propose method (Eq. (5)). In contrast to results obtained by the conventional method, the boundaries of both Mother an Daughter are preserved quite well.

4. Conclusions

In this paper, we propose a new algorithm for semi automatic segmentation using region growing. We define new difference measure for the region growing algorith which exploits features of color information efficiently an produces improved segmentation results that are suitable fo content-based video coding, such as MPEG-4.

Acknowledgments

This work was supported in part by the Korea Scienc and Engineering Foundation (KOSEF) through the Ultra Fast Fiber-Optic Networks (UFON) Research Center a Kwangju Institute of Science and Technology (K-JIST).

References

[1] T. Aach and A. Kaup, "Statistical model-based chang detection in moving video," Signal Processing, vol. 3 pp. 165-180, 1993.

[2] M. Hotter and R. Thoma, "Image segmentation based o object oriented mapping parameter estimation," Sign Processing, vol. 15, pp. 315-334, 1988.

[3] P. Salembier, "Morphological multiscale segmentati for image coding," Signal Processing, vol. 38, pp. 35 386, 1994.

[4] L. Vincent and P. Soille, "Watersheds in digital space an efficient algorithm based on immersion simulations IEEE Trans. Pattern Analysis and Machine Intelligenc vol. 13, pp. 583-598, June 1991.

[5] J. Choi, M. Kim, M. Lee and C. Ahn, "Automatic se mentation based on spatio-temporal information ISO/IEC JTC1/SC29/WG11, MPEG4/m2091, April 19

[6] J. Choi, M. Kim, M. Lee and C. Ahn, "A user-assist segmentation method for video object plane generatior ITC-CSCC' 98, pp. 7-10, July 1998.

[7] M. Kim and Y. Ho, "Semi-Automatic Segmentation A gorithm Using Double Labeling Mehtod," IEEE Regi 10 Conference, pp. 746-749, September 1999.