SEMI-AUTOMATIC SEGMENTATION BY A DOUBLE LABELING METHOD

Minho Kim and Yo-Sung Ho

Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-Dong Puk-Gu, Kwangju, Korea
E-mail: minokim@geguri.kjist.ac.kr, hoyo@kjist.ac.kr

Abstract

This paper proposes a semi-automatic segmentation algorithm using a double labeling method. Semi-automatic segmentation consists of intra-frame segmentation and inter-frame segmentation. For intra-frame segmentation, we obtain a globally labeled mask from the boundary information provided by the user and a locally labeled mask from the intensity information of the image itself. By merging the globally and locally labeled masks, we complete the intra-frame segmentation operation. Inter-frame segmentation is accomplished by tracking the video objects (VOs) that are defined from intra-frame segmentation. Semi-automatic segmentation using the double labeling method resolves the object correspondence problem, which is inherent in automatic segmentation.

I. INTRODUCTION

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

For this purpose, several automatic segmentation algorithms have been developed. Typical algorithms include change detection-based segmentation [1,2], motion-based segmentation [6], morphological segmentation [7,8] and spatio-temporal segmentation [3,5].

Change detection-based segmentation uses 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 the uncovered background from video objects.

Motion-based segmentation [6,9] solved the covered and uncovered background problem by assuming that an 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 approaches utilize temporal information by taking differences between two successive picture frames, morphological segmentation employs mathematical morphology operations which are very useful for the segmentation purpose. Mathematical morphology can efficiently deal with geometrical features, such as size, shape, contrast and connectivity, which can be considered as object-oriented features. Thus, the morphological segmentation algorithm can preserve the boundaries of video objects. However, each partitioned region itself may not represent a meaningful video object.

There have been several attempts to solve the segmentation problems mentioned above by exploiting both temporal and spatial information [3,5]. However, they could not solve the problems completely, because the definition of video objects is very subjective and automatic segmentation methods use only the 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.

In this paper, we propose a new semi-automatic segmentation algorithm using a double labeling method. Two key ideas are (1) combining the boundary information of the user and the intensity information of the image efficiently, and (2) tracking the video objects by motion prediction.

II. SEMI-AUTOMATIC SEGMENTATION

As shown in Figure 1, semi-automatic segmentation consists of intra-frame segmentation and inter-frame segmentation. In the first picture frame, we extract video objects of interest by morphological segmentation tools. Then, we apply inter-frame segmentation for the subsequent picture frames until we encounter one of the following three events: (1) appearance of new video objects, (2) scene changes, or (3) unsatisfactory segmentation results. If any of the three events occurs, we restart the intra-frame segmentation operation.

A. 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. As shown in Figure 2, intra-frame segmentation is composed of a global labeling, a local labeling, and merging the global and the local labeling.

1. Global Labeling by User Interaction

For the global labeling, the user can provide the boundary information of video objects. User interaction can have a form of changing segmentation parameters for each picture frame, selecting the region of interest, or

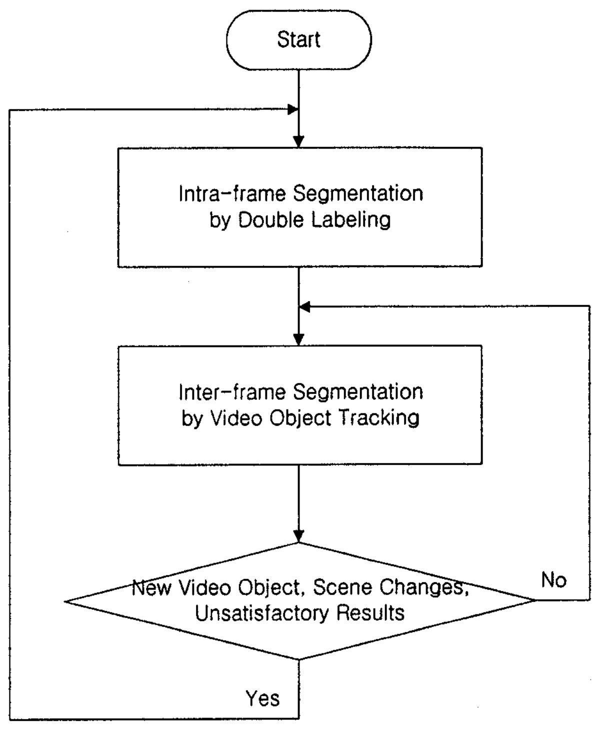


Figure 1: Major Operations of Semi-Automatic Segmentation

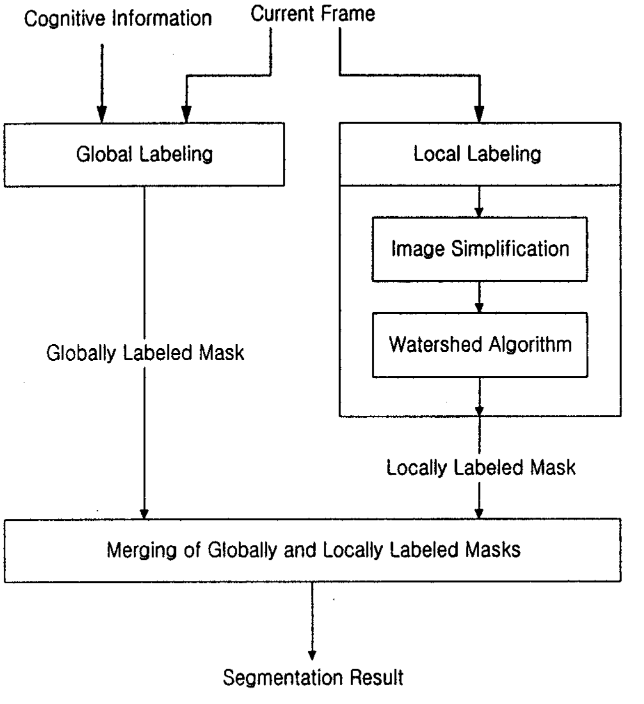


Figure 2: Intra-Frame Segmentation

indicating video objects of interest by drawing the uncertainty region around boundaries of the video objects. In this paper, we use the last option to obtain a globally labeled mask. This globally labeled mask contains the background, video objects and uncertainty areas.

2. 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 guarantees a good preservation of the boundaries of the video objects. However, other region labeling algorithms need some adjustments of segmentation parameters, such as k values in the k-means algorithm, to produce reasonable segmentation results for various kinds of images.

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 kind of mosaic. Thus, we call the result produced by morphological segmentation as a locally labeled mask.

The morphological segmentation algorithm [7] contains three main operations: (1) image simplification, (2) marker extraction, and (3) region decision.

Image simplification is accomplished by the morphological filters, especially by the 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;

$$g = \delta(f) - \varepsilon(f) \qquad (1)$$

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.

Then, we carry out the region decision operation using the watershed algorithm [8]. 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.

3. Merging of Globally and Locally Labeled Masks

Figure 3 illustrates the merging operation of the 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, the boundaries of video objects is well preserved, because the image is partitioned by the topographic property, i.e., the characteristics of the image obtained by 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.

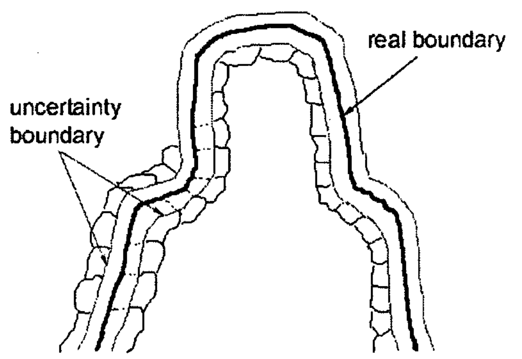


Figure 3: Merging of Globally and Locally Labeled Mask

Figure 3 is produced by overlapping the locally labeled mask in the globally mask. In Figure 3, the thick line represents the actual boundary of the video object, and the space between the lines, marked as uncertainty boundary, represents the uncertainty areas of the globally labeled mask. Note that there exist some regions that partially belong to the uncertainty region and partially belong to the video object or the background, which is originally the same region in the locally labeled mask. Thus, we can assign this uncertainty region to the video object or the background.

With this operation, neighboring parts of video objects or the background grows into uncertainty areas, respectively. The real boundary is located in the point where the video object and the background meet.

After the merging operation, there can exist the locally labeled regions which are totally included in the uncertainty area. If this case happens, we remove these regions by merging them into a neighboring region that has the best similarity, specifically the region that has the smallest difference in terms of the region average. Finally, we obtain the mask for intra-frame segmentation.

B. Inter-Frame Segmentation

Inter-frame segmentation extracts video objects from the background by tracking the video objects of the previous frame, as explained in Figure 4. Inter-frame segmentation consists of object prediction followed by boundary fitting. For the object prediction, we project each video object of previous frame into the current frame by the 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.

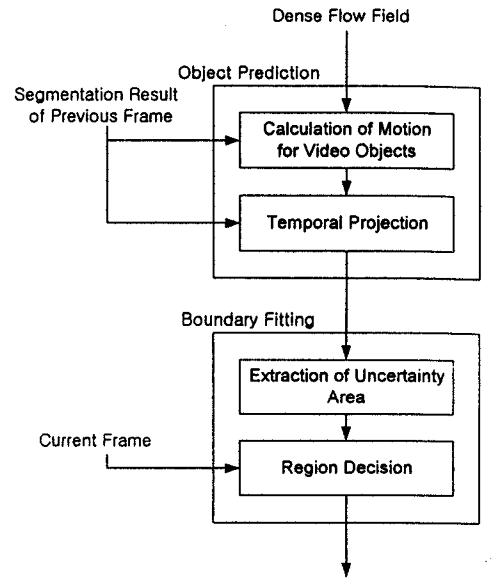


Figure 4: Inter-Frame Segmentation

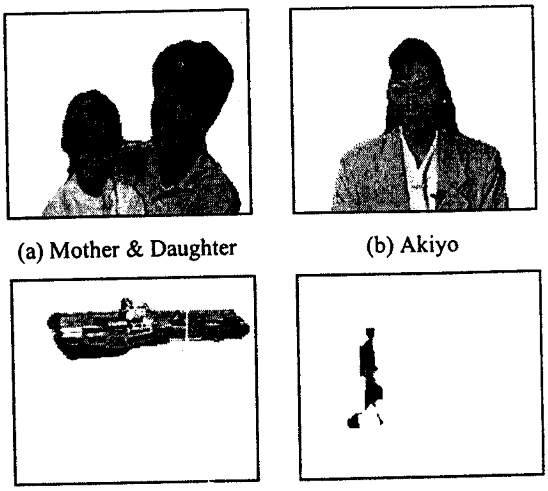
Since object prediction usually creates error regions due to inexact motion estimation and occlusion, we should perform boundary fitting. For boundary fitting, we extract an uncertainty area for neighboring pixels of the video object boundaries. In addition, we mark the unlabeled (uncovered) or multi-labeled (overlapped) pixels that are produced in the temporal projection operation as an uncertainty area. Then, we find the actual boundary by the region growing algorithm, which makes use of the brightness information of the current picture frame [4].

III. EXPERIMENTAL RESULTS

In this section, we compare simulation results of typical automatic segmentation and the proposed semi-automatic segmentation. For subjective evaluation, we have selected MPEG-4 test sequences: "Mother & Daughter (MD)", "Akiyo", "Container Ship (CS)", and "Hall & Monitor (HM)" sequences, which have the QCIF format of 176x144 pixels.

Simulation results of the semi-automatic segmentation algorithm are shown in Figure 5, where we can see that we obtain video objects of interest with exact boundaries. In Figure 6, we compare results of the semi-automatic segmentation (SAS) algorithm and automatic segmentation algorithms: change detection-based segmentation (CDBS), motion-based segmentation (MBS), and spatio-temporal segmentation (STS). As we can see, the semi-automatic segmentation algorithm produces superior results to the automatic segmentation algorithms.

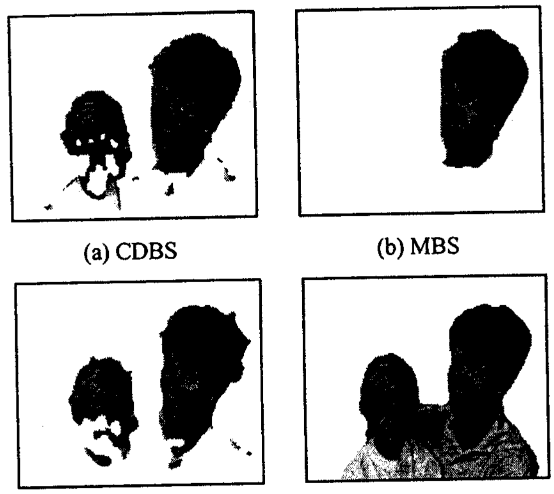
For objective evaluation, we define a distortion as the sum of pixel differences between the segmented data provided by MPEG and various segmentation masks. Figure 7 and Figure 8 indicate that the semi-automatic segmentation algorithm has smaller distortion than automatic segmentation algorithms



(a) Mother & Daughter (b) Akiyo

(c) Container Ship (d) Hall & Monitor

Figure 5: Results of Intra-Frame Segmentation



(a) CDBS (b) MBS

(c) STS (d) SAS

Figure 6: Subjective Comparison

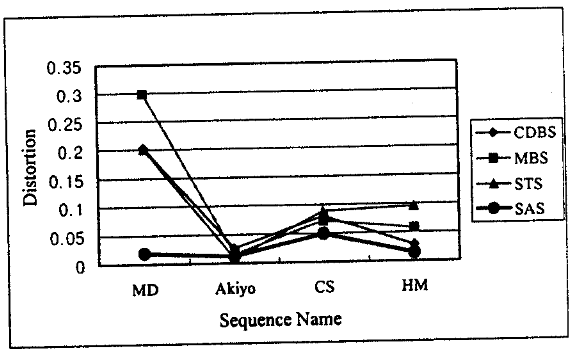


Figure 7: Objective Comparison (Intra-Frame)

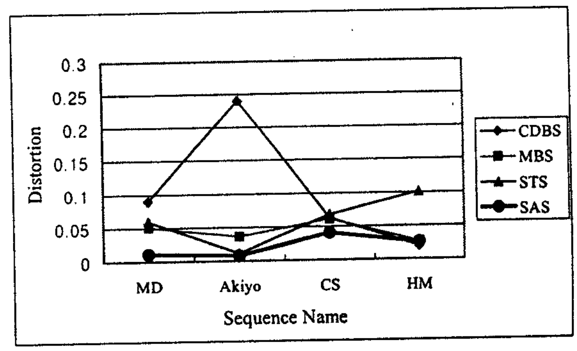


Figure 8: Objective Comparison (Inter-Frame)

IV. CONCLUSIONS

In this paper, we propose a new semi-automatic segmentation algorithm using a double labeling method. By combining the boundary information provided by the user with the intensity information of the image itself, the semi-automatic segmentation algorithm produces improved segmentation results which are suitable for content-based video coding, such as MPEG-4.

ACKNOWLEDGMENT

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

REFERENCES

[1] T. Aach and A. Kaup, "Statistical model-based change detection in moving video", *Signal Processing*, vol. 31, pp. 165-180, 1993.

[2] T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random field", *Signal Processing: Image Communnication*, vol. 7, pp. 147-160, 1995.

[3] J.G. Choi, S.W. Lee and S.D. Kim, "Spatio-temporal video segmentation using a joint similarity measure", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 297-286, April 1997.

[4] J.G. Choi, et al., "A user-assisted segmentation method for video object plane generation", *ITC-CSCC' 98*, pp. 7-10, July 1998.

[5] J.G. Choi, M.C. Kim, M.H. Lee and C.D. Ahn, "Automatic segmentation based on spatio-temporal information", *ISO/IEC JTC1/SC29/WG11, MPEG4/m2091*, April 1997.

[6] M. Hotter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, vol. 15, pp. 315-334, 1988.

[7] P. Salembier, "Morphological multiscale segmentation for image coding", *Signal Processing*, vol. 38, pp. 359-386, 1994.

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

[9] N. Diehl, "Object-oriented motion estimation and segmentation in Image Sequences", *Signal Processing: Image Communication*, vol. 3, pp. 23-56, 1991.