Video Matting based on Background Estimation

J.-H. Moon, D.-O Kim, and R.-H. Park

Abstract—This paper presents a video matting method, which extracts the foreground and alpha matte from a video sequence. The objective of video matting is finding the foreground and compositing it with the background that is different from the one in the original image. By finding the motion vectors (MVs) using a sliced block matching algorithm (SBMA), we can extract moving regions from the video sequence under the assumption that the foreground is moving and the background is stationary. In practice, foreground areas are not moving through all frames in an image sequence, thus we accumulate moving regions through the image sequence. The boundaries of moving regions are found by Canny edge detector and the foreground region is separated in each frame of the sequence. Remaining regions are defined as background regions. Extracted backgrounds in each frame are combined and reframed as an integrated single background. Based on the estimated background, we compute the frame difference (FD) of each frame. Regions with the FD larger than the threshold are defined as foreground regions, boundaries of foreground regions are defined as unknown regions and the rest of regions are defined as backgrounds. Segmentation information that classifies an image into foreground, background, and unknown regions is called a trimap. Matting process can extract an alpha matte in the unknown region using pixel information in foreground and background regions, and estimate the values of foreground and background pixels in unknown regions. The proposed video matting approach is adaptive and convenient to extract a foreground automatically and to composite a foreground with a background that is different from the original background.

Keywords— Background estimation, Object segmentation, Block matching algorithm, Video matting.

I. INTRODUCTION

AS demands on entertainment such as movies and games are increasing, image composition is an important issue that changes images more interesting and makes impossible situations in an actual world real. For this objective, object segmentation techniques using edge information to extract the foreground have been presented [1]. But simply cutting the foreground region cannot be directly used for image composition, because boundary pixels represent the blended region with foreground and background pixels. To reduce blended regions in image composition, matting has been commonly used.

Blue screen matting has been commonly used to extract the interesting foreground, in which a photograph of the foreground against a constant colored background such as a blue or green screen is used. Blue screen matting is easy to implement, thus has been used very widely in TV broadcasting programs and movies. But this method suffers from maintaining a constant background due to shadow of the foreground or light direction. To overcome this limitation, a method was presented in which two photographs against two different screens were taken [2]. However this method is not suitable for video, because making two video sequences with the same foreground is impossible and impractical. It is only suitable for the stationary object that is not moving at all.

Bayesian matting gives good results for still images [3], however it needs user interaction called a trimap that indicates three types of regions: "foreground", "background", and "unknown" regions. It is suitable for still images. However this user interaction is boring and time-consuming. So a video matting method was proposed [4], in which initial user interaction is needed and a trimap is interpolated using optical flow that keeps tracks of foreground, background, and unknown regions of the previous frame, also requiring user interaction.

Automatic video segmentation has been investigated in the context of content-based coding such as moving picture experts group (MPEG)-4 [5]. It estimates optical flow and characterizes independently moving components using morphological filtering. However it has a difficulty in defining boundary pixel values such as hair, if foreground boundaries are not deliberately cut.

This paper proposes a video matting method that extracts the foreground and alpha matte from a video sequence. It consists of three steps: sliced block matching algorithm (SBMA), background estimation, and Bayesian matting. First we compute the current frame difference (FD) of a block and employ the SBMA [6] for motion estimation. The SBMA yields accurate motion vectors (MVs) in object boundary blocks, in which up to three different objects with different motions are assumed. Based on the computed MVs of the block, moving regions are obtained. Edge detection in moving regions gives the foreground region. In each frame, the background behind the object will be exposed and the exposed background is accumulated as a background. By accumulating uncovered backgrounds, an integrated background is estimated. With the estimated background, we compare two successive frames in a video sequence; if the pixel does not have a background value (i.e., if the gray level difference from that of the background is smaller than the threshold) the pixel is defined as the pixel in

J.-H. Moon is with the Department of Electronic Engineering, Sogang University, Seoul, Korea (phone: +82 2 716 4514; fax: +82 2 716 4514; e-mail: jhmoon@sogang.ac.kr).

D.-O Kim is with the Department of Electronic Engineering, Sogang University, Seoul, Korea (e-mail: kdo@eevision1.sogang.ac.kr).

R.-H. Park is with the Department of Electronic Engineering, Sogang University, Seoul, Korea (e-mail: rhpark@sogang.ac.kr).

World Academy of Science, Engineering and Technology
International Journal of Aerospace and Mechanical Engineering
Vol:1, No:2, 2007

the foreground. Remaining pixels are defined as unknown regions that need more computations. After finding moving area, by comparing adjacent image frames in a video sequence, background regions are tracked under the assumption that the foreground is moving and the background is stationary.

To extract unknown regions that mostly contain object boundaries, Bayesian matting is employed [2]. Segmentation information that classifies an image into foreground, background, and unknown regions is called a trimap. Matting process extracts an alpha matte in the unknown region using pixel information of foreground/background regions and estimates foreground/background pixel values. Finally, using foreground and alpha matte values, we can update the background.

In summary, in the proposed video matting algorithm, moving regions are detected using the SBMA and the background is estimated using edge detection. The separated background in each frame forms a single integrated background. Bayesian matting uses information about foreground/background regions, estimates foreground/background values in the unknown region, and extracts an alpha matte in boundaries. The proposed method generates a trimap using the FD and Bayesian matting. Simulation results show that the proposed video matting algorithm can faithfully extract boundaries with low gradient values. Moreover, it is adaptive to extract a foreground automatically and to composite a foreground with a background.

II. VIDEO MATTING

A. SBMA

We first split an image into a number of non-overlapping blocks and classify each block into three regions (P, N, and Z regions) using the FD. The FD is defined as:

$$FD(i, j) = I_t(i, j) - I_{t-1}(i, j) \qquad (1)$$

where $I_t(i, j)$ represents the pixel intensity at location (i,j) in the tth image. Then three regions are separated as:

$$
\begin{aligned}
P &= \{(i, j) \,|\, FD\,(i, j) > Th\,,\ 0 \le i < N_1,\ 0 \le j < N_2\} \\
N &= \{(i, j) \,|\, FD\,(i, j) < -Th\,,\ 0 \le i < N_1,\ 0 \le j < N_2\} \\
Z &= \{(i, j) \,|\, |FD(i, j)| < Th,\ 0 \le i < N_1,\ 0 \le j < N_2\}
\end{aligned}
$$

$$\qquad (2)$$

where N_1 and N_2 denote the numbers of pixels along the horizontal and vertical directions, respectively, and Th represents a selected threshold. Local MVs of each region are estimated by the block matching algorithm (BMA), with the assumption that each block has multiple (up to three) MVs. The SBMA yields accurate MVs in object boundary blocks.
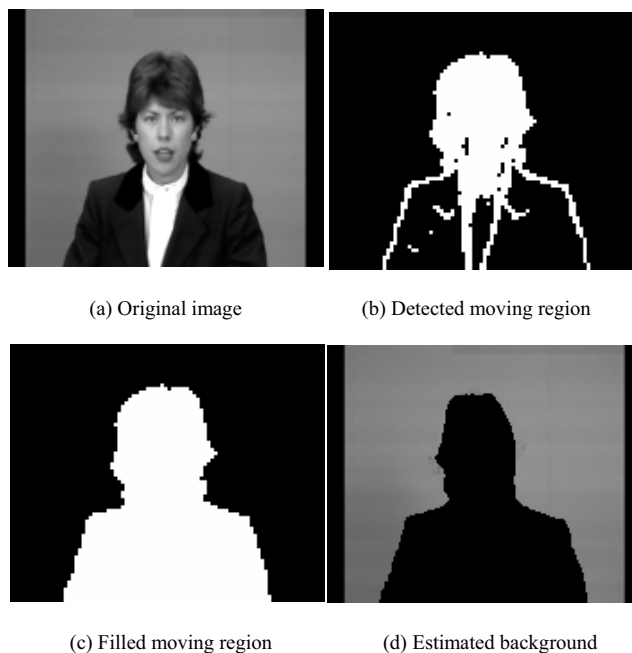

(a) Original image   (b) Detected moving region

(c) Filled moving region   (d) Estimated background

Fig. 1: Background estimation.

B. Background Estimation

By SBMA, moving regions can be detected. Moving regions do not include all of foregrounds because foregrounds do not necessarily move through all image frames. So, moving regions are accumulated through the image sequence. They include an entire foreground and some backgrounds. To separate foregrounds, we employ Canny edge detector, by which foreground and background regions can be segmented. Segmented backgrounds in each frame are accumulated and finally an integrated single background is estimated. This estimated single background is used to generate a trimap that yields accuracy Bayesian matting. Fig. 1 shows background estimation procedure. Fig. 1(a) shows an original image and Fig. 1(b) illustrates a detected moving region (white region). Fig. 1(c) shows a filled moving region from Fig. 1(b) and Fig 1(d) shows the estimated background.

C. Trimap Generation

After a single background is obtained, each frame is compared with a single background. By background subtraction, foreground regions are obtained. Boundaries of foreground regions are assigned as unknown regions by employing an $n \times n$ window. The interior (exterior) of unknown regions is defined as the foreground (background). This information is called a trimap.

D. Bayesian Matting

Most images are acquired by charge coupled device (CCD) cameras. The pixel value is represented by the integrated light intensity over the corresponding region. Pixel values in borders
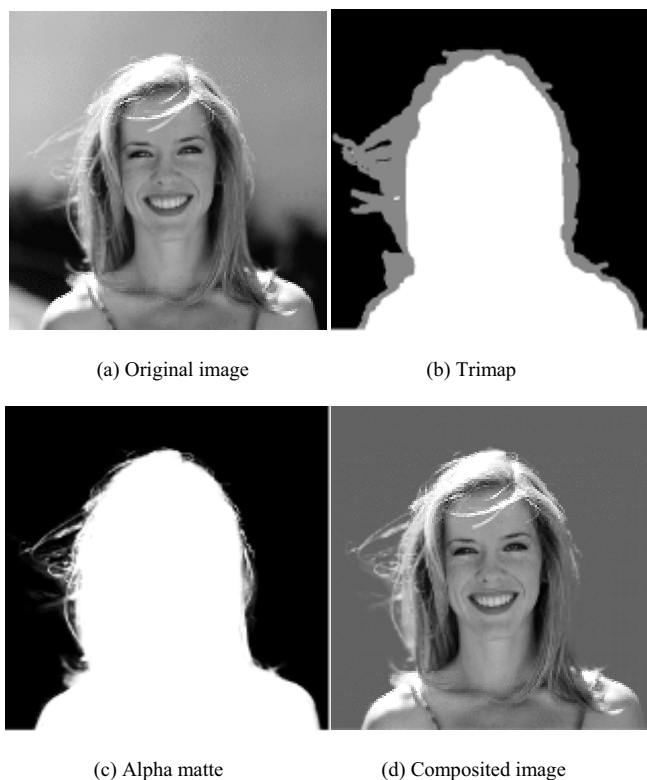
World Academy of Science, Engineering and Technology
International Journal of Aerospace and Mechanical Engineering
Vol:1, No:2, 2007

(a) Original image   (b) Trimap

(c) Alpha matte   (d) Composited image

Fig. 2: Bayesian matting.

Foreground, Background, Alpha matte

Fig. 3: Flowchart of the proposed video matting algorithm.

between objects or backgrounds are compositions of different object colors. A compositing equation expressed as

$$C = \alpha F + (1-\alpha)B \qquad (3)$$

was proposed [7], where C, F, and B signify the observation, foreground, and background colors at a pixel, respectively, with $0 \le \alpha \le 1$ representing the composition weight of foreground and background pixels.

A Bayesian approach to digital matting [3] uses *maximum a posteriori* (MAP) method to solve (3). When the observed color C at pixel is given, using the Bayes's rule, the proper foreground color, background color, and opacity are obtained by maximizing the sum of log likehoods:

$$\begin{aligned} &\arg\max_{F,B,\alpha} \ P(F,B,\alpha|C) \\ &= \arg\max_{F,B,\alpha} P(C|F,B,\alpha)P(F)P(B)P(\alpha)/P(C) \\ &= \arg\max_{F,B,\alpha} L(C|F,B,\alpha) + L(F) + L(B) + L(\alpha), \end{aligned} \qquad (4)$$

where $L(.)$ denotes the log likelihood defined by $L(\cdot) = \log P(\cdot)$, with $P(\cdot)$ denoting the probability. In (4), $P(C)$ is dropped because it is a constant. To perform this matting procedure, a trimap that indicates foreground, background, and unknown regions is needed. Fig. 2 shows a procedure of Bayesian matting. Fig. 2(a) shows an original image and Fig. 2(b) illustrates a trimap obtained by user interaction containing up
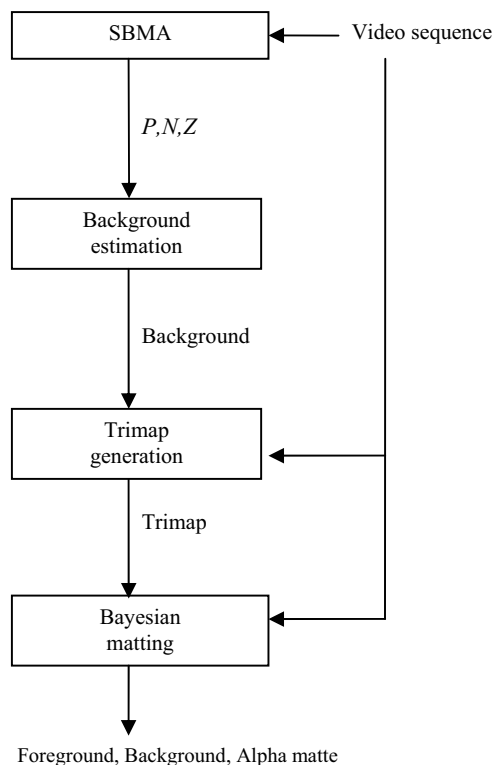
to three types of regions (foreground, background, and unknown regions). Fig. 2(c) shows an alpha matte that represents opacity and Fig 2(d) shows an output image in which the foreground is composited with a constant color background.

E. Video Matting

Fig. 3 shows a flowchart of the proposed video matting procedure, consisting of four steps: SBMA, background estimation, trimap generation, and Bayesian matting. In the first step, we compute MVs of an image sequence, extracting moving regions. Using motion estimation to find moving area, rather than using the FD, avoids detecting incorrect moving regions caused by practical noise and illumination changes. We assume that the foreground is moving, whereas the background is stationary. But we cannot assure that the entire foreground is moving. To find an integrated foreground region, moving regions through the image sequence are combined. Also to separate the foreground from the background, edges in moving regions are utilized. The separated background is accumulated through the whole image sequence, yielding an integrated background. By comparing each frame in a sequence with the estimated background, the foreground region is obtained. To define unknown regions, we employ an $n \times n$ window and contour boundary information between the foreground and background. Then the trimap is obtained. The next step is a Bayesian matting procedure. Bayesian matting extracts an alpha matte using a trimap, from which we can assign opacity

World Academy of Science, Engineering and Technology
International Journal of Aerospace and Mechanical Engineering
Vol:1, No:2, 2007

and finally extract the foreground and alpha matte. We composite a new background and foreground in each frame of the video sequence with the extracted alpha matte.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

Fig. 4 shows video matting results. Fig. 4(a) shows an original image and Fig. 4(b) shows the estimated background. Figs. 4(c) and 4(d) illustrate a trimap and alpha matte, respectively. Fig. 4(e) shows the result of the proposed video matting algorithm, in which the composited foreground looks naturally. The proposed video matting uses an alpha matte to composite the foreground with a different background. The alpha matte can make compositing with other background more naturally. Object segmentation using edge detector alone or simple cutting based on the gradient information can separate objects also. But these methods suffer from compositing the different backgrounds because pixels in object boundaries include background pixel intensity and color. The proposed video matting method with Bayesian matting successfully removes background color and composites a new background color using an alpha matte.

IV. CONCLUSIONS

In this paper, we propose a video matting algorithm based on background estimation using SBMA and edge detection. Using the SBMA, we find moving regions. By accumulating backgrounds with edge information, a single integrated background is obtained. Using a single background through the image sequence, a trimap is obtained. Finally Bayesian matting extracts an alpha matte and the foreground region. The proposed video matting algorithm is efficient for image composition and can be applied to entertainment applications such as movies and games.

REFERENCES

[1] E. N. Mortensen and W. A. Barrett, "Intelligent scissor for image composition," in *Proc. SIGGRAPH 1995*, pp. 191-198, Los Angeles, CA, USA, Aug. 1995.
[2] A. R. Smith and J. F. Blinn, "Blue screen matting," in *Proc. SIGGRAPH 1996*, pp. 259-268, New Orleans, LA, USA, Aug. 1996.
[3] Y.-Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. Computer Vision and Pattern Recognition*, vol. 2, pp. 264-271, Kauai, Hawaii, USA, Dec. 2001.
[4] Y.-Y. Chuang, A. Agarwala, B. Curless, D. Salesin, and R. Szeliski, "Video matting of complex scenes," in *Proc. SIGGRAPH 2002*, pp. 243-248, San Antonio, TX, USA, July 2002.
[5] T. Meier and K. N. Ngan, "Video segmentation for content-based coding," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 9, pp. 1190-1203, Dec. 1999.
[6] E. D. Sciascio and C. Guarangnella, "Object oriented motion estimation by sliced-block matching algorithm," in *Proc. Int. Conf. Pattern Recognition*, Barcelona, Spain, vol. 3, pp. 3865-3868, Sep. 2000.
[7] T. Porter and T. Duff, "Compositing digital images," in *Proc. SIGGRAPH 1984*, vol. 18, no. 3, pp. 253-259, Minneapolis, MN, USA, July 1984.

(a) Original image

(b) Estimated background

(c) Trimap

(d) Alpha matte

(e) Composited image

Fig. 4: Video matting procedure.