MULTI-VIEW IMAGE MATTING AND COMPOSITING USING TRIMAP SHARING FOR NATURAL 3-D SCENE GENERATION

Myung-Han Hyun, Sung-Yeol Kim, and Yo-Sung Ho

Gwanju Institute of Science and Technology (GIST), Korea

ABSTRACT

Digital matting for extracting foreground objects from an image is an important process to generate special effects in the movie industry and the broadcasting center. Recently, a digital matting algorithm has been developed to create an alpha matte using a well-focused image generated from multi-view images. However, this method could generate only a single-view alpha matte, even though it employed multiple cameras. In this paper, we propose a new estimation scheme for multi-view alpha mattes by sharing the trimap of the reference view. After we extract foreground objects from all view images, we composite the foreground objects with the corresponding background images captured from the same multi-view camera arrangement. Experimental results demonstrate that multi-view composite images can generate reasonably natural 3-D views through the stereoscopic monitor.

Index Terms— digital matting, multi-view image matting, multi-view camera system, trimap sharing

1. INTRODUCTION

Efficient image and video compositing techniques are required to make special effects in the movie industry or the broadcasting center. In general, a composite image is divided into two layers: foreground object and background. To extract a foreground object, referred to as a *foreground matte*, we remove the background of the original image by considering an alpha value α that represents the pixel opacity of the image. This technique is known as digital matting [1]. Meanwhile, digital compositing is to combine a foreground matte with an arbitrary background by using the alpha value α [2]. We can blend both foreground and background pixels by an equation $I = \alpha F + (1-\alpha)B$, where I, F and B are the composite, foreground and background images, respectively.

Blue screen matting is widely used for digital matting [3]. Since the blue screen matting algorithm uses monotonous blue or green backgrounds, it is easy to extract a foreground object from them. However, if the foreground object contains the background constraint color, it is hard to pull out the foreground object efficiently. Furthermore, a virtual studio environment is required for the blue screen matting. In order to overcome the limitation of blue screen matting, natural image matting, which has fewer constraints for backgrounds, has been actively studied in the field of computer vision [4]. However, the natural image matting requires user assistances and more complicated algorithms than the blue screen matting. Moreover, we need to make strenuous efforts to extract the foreground object from complex background scenes [5]. To obtain the enhanced alpha matte in the natural image matting, we can also use secondary operations by an image gradient [6]. Although the previous works enable us to represent complex boundaries correctly, they usually take too tedious operations.

Recently, a digital matting algorithm using multi-view cameras has been developed to create an accurate alpha matte [7]. It can generate an alpha matte fast and automatically. However, even though this work employs a multi-view camera system, it can only generate a single-view alpha matte, because it assumed that the alpha value α is fixed in all view images. As a result, the work cannot produce multi-view composite images from the multi-view camera system. In addition, since the previous work suffers from a large amount of ad-hoc operations to create a trimap, such as double-thresholding and the selection of a structuring element, the unknown region of the trimap is tended to be isolated and broadened. Therefore, it is difficult to estimate the alpha value in the unknown region.

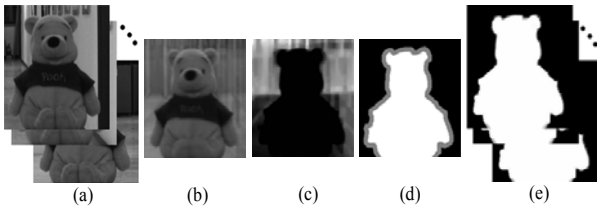
In this paper, we propose a new digital matting algorithm to estimate multi-view alpha mattes using a multi-view camera system. The main contribution of our work is that we first propose the concept and methodology of the multi-view image matting and compositing. In this work, we generate view-dependent alpha mattes to extract each foreground object from multi-view images by sharing a trimap of a reference view. Furthermore, we consider multi-view background images captured from the identical camera system for digital compositing. Thus, we can generate 3-D scenes using multi-view composite images. We can also reduce the overall processing time in comparison with the conventional matting methods that independently extracts the foreground object from each camera.

2. MULTI-VIEW IMAGE MATTING AND COMPOSITING

We need a 3-D image editing system to generate multi-view composite images. Figure 1 shows the proposed procedure for the multi-view image matting and compositing. First, we film multi-view images using a multi-view camera system as shown in Fig. 1(a). Second, we make a synthetic aperture image (SAI) by moving other view images to the center view image, i.e., the reference view, of the multi-view camera system [7]. After creating SAI, we convert it into a variance image by calculating a variance of its corresponding pixels. Third, in order to generate a trimap in Fig. 1(d), we apply dilation and erosion operations into a binary image generated by the variance image. The trimap contains the foreground object, background and unknown areas. Fourth, the

978-1-4244-1755-1/08/$25.00 ©2008 IEEE

trimap is shared with other views and used to create multi-view alpha mattes by estimating the opacity of unknown areas. Finally, we extract and composite the multi-view foreground objects from multi-view images using multi-view alpha mattes in Fig. 1(e).



(a) (b) (c) (d) (e)

Fig. 1. Matting procedure. (a) Multi-view images, (b) SAI, (c) Variance image, (d) Trimap, and (e) Multi-view alpha mattes.

2.1. Synthetic Aperture Image

In order to generate SAI, we overlap corresponding image pixels in each camera by shifting the disparity between the foreground object of the reference view and foreground object of the other camera [8]. As shown in Fig. 1(b), SAI is synthetically refocused according to the foreground object's depth plane. The characteristic of SAI is that it is well-focused on the foreground object and blurred out at the background. In our work, we manually pick the foreground depth by interactively sliding the synthetic plane of focus through the scene. From SAI, we compute variance statistics to get a variance image, as shown in Fig. 1(c).

2.2. Trimap Generation and Sharing

In the previous method, the trimap can be generated by taking two threshold values T_1 and T_2. Especially, it is difficult to determine T_2 because it varies in a broad range from 1000 to 5000. Basically, the threshold T_2 depends on the number of images and the characteristic of images. Furthermore, two threshold values are selected from ad-hoc and subjective operations by trial and error.

In order to make a trimap automatically, we adapt automatic threshold algorithm to make the binary image from the variance image. In the proposed method, we first calculate the histogram of the variance pixel values, and set an initial threshold T as 20 because most of pixel intensities of the variance image are skewed around the lower intensity 20 in the histogram. Then, we separate the histogram into two areas using the initial threshold T and calculate the mean intensity of each area. Finally, we update the threshold T as the average of each mean, and repeat the averaging and updating procedures until T is converged. As a result, we can make a binary image automatically.

From the binary image, we make a trimap by dilating outward and eroding inward to specify the unknown area. To this end, the foreground object, background and unknown areas are represented by

$$Foreground \quad : \mathrm{A} = \{\mathbf{I} | \mathbf{var}(\mathbf{I}) < T\} \tag{1}$$

$$Background \quad : \mathrm{B} = \{\mathbf{I} | \mathbf{var}(\mathbf{I}) \geq \mathrm{T}\} \tag{2}$$

$$Unknown \quad : \mathrm{C} = \{(\mathrm{A} \oplus \mathrm{Z}) \boxplus (\mathrm{A} \ominus \mathrm{Z})\}^{\mathrm{c}} \tag{3}$$

The foreground set A contains the region in which the variance value is less than T, and the background set B contains the region in which the variance value is greater than T. The set A, set B, and set C, which represent the trimap, contain pixel intensity values 255, 0, and 128, respectively. The Symbols $\oplus$, $\ominus$, and $\boxplus$ are morphological operations and represent dilation, erosion and exclusive-or, respectively. var(I) is the pixel intensity of the variance image at the pixel position I.

After creating the trimap, we need to share it for other view images to generate multi-view alpha mattes. In order to determine the size of the unknown area in the trimap, we use the concept of 3-D warping techniques. After selecting the two points on the foreground in the reference view, whose distance is maximized, we perform 3-D warping into the corresponding location of the right most and left most views. Then, we calculate the distance between the warped two points. Likewise, we set the size of the unknown area as width in Eq. (4) and Eq. (5).

$$Width = \text{longest distance} - \text{shortest distance} \tag{4}$$

$$Height = 1/2 \times Width \tag{5}$$

Ideally, we do not care the disparity in the vertical direction because we assume that the height of each camera from the ground is identical. However, we should consider the vertical disparity due to the possibility of vertical height difference in actual application. Here, we set the vertical size of the unknown areas as a half of width.

2.3. View-dependent Alpha Matte Generation

In other to generate view-dependent alpha mattes with the shared trimp, we first convert the RGB color space for the multi-view images into YCbCr color space, and apply Gaussian filtering into Cb components to reduce the noise around the foreground boundaries. Then, we extract and label edges from unknown areas using the canny edge algorithm. Finally, we make the contour of the foreground by connecting the labeled edges to make multi-view alpha mattes.

However, during view-dependent alpha matte generation, we cannot obtain robust edges from the canny edge algorithm. To overcome the problem in the edge extraction, we apply the histogram equalization operator into the multi-view images. Then, we again apply the canny edge operation with Cb components of the histogram equalized multi-view images. Finally, we merge the extracted edges of original multi-view images with the extracted edges of histogram equalized multi-view images.

3. EXPERIMENTAL RESULTS

In order to evaluate the proposed multi-view image matting, we used a 1-D parallel multi-view camera system where seven Multi Sync IEEE-1394b cameras were equipped with the camera baseline of 5cm. Figure 2 shows the multi-view camera system.

Fig. 2. Multi-view camera system

Figure 3 and Figure 4 show the results of multi-view matting and compositing. As shown in Fig. 3 and Fig. 4, we extracted foreground objects of each view and composited them with corresponding multi-view background images from the same camera environment. The previous method had no contribution for the 3-D scene generation because it only considered the single-view foreground on an arbitrary background image.

(a) Multi-view image of "Pooh"

(b) Multi-view alpha matte of "Pooh"

(c) Multi-view foreground of "Pooh"

(d) Multi-view image of "Person"

(e) Multi-view alpha matte of "Person"

(f) Multi-view foreground of "Person"

Fig. 3. Results of the "Pooh" and "Person" test sequences

Fig. 4. Results of multi-view composite images

However, the proposed method considered the multi-view foreground objects and the corresponding multi-view background images captured by the same multi-view camera system. As a result, as shown in Fig. 4, multi-view composite images generated by the proposed method could be used as a multi-view video.

In order to prove the multi-view composite images can generate the multi-view video, we displayed the two adjacent images of the multi-view composite images with a stereoscopic monitor as shown in Fig. 5. We also displayed the stereoscopic images captured by the two adjacent cameras in the multi-view camera system with the stereoscopic monitor. We could experience natural 3-D effects by not only the stereoscopic images but also stereoscopic composite images. Thus, the proposed multi-view image matting can be used for a 3-D image editing system.

(a) Stereo input (b) Stereoscopic view

Fig. 5. Results of the stereoscopic view

Figure 6 and Figure 7 show the error rate of the two test images, "pooh'' and "person", using the previous and proposed methods. We used error evaluation method from Eq. (6) and Eq. (7),

$$E_U = \frac{\mathrm{M} - (\mathrm{M} \cap \mathrm{N})}{S_M}, E_O = \frac{\mathrm{N} - (\mathrm{M} \cap \mathrm{N})}{S_N} \quad (6)$$

$$Error\ Rate = (E_U + E_O) \times 100, \quad (7)$$

where M is pixel number of ground truth image from a image tool, and N is pixel number of foreground object from proposed method. The term S_M and S_N are the numbers of M and *N*. E_U is the error rate of under-extraction and E_O is the error rate of over-extraction. With under- and over- extraction, we mean that the corresponding area is and is not to be extracted. As you can see from the quantitative error results, our method provides the lower error rate than the previous method.

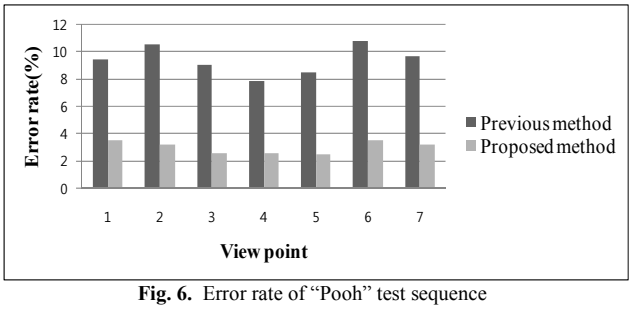


Fig. 6. Error rate of "Pooh" test sequence

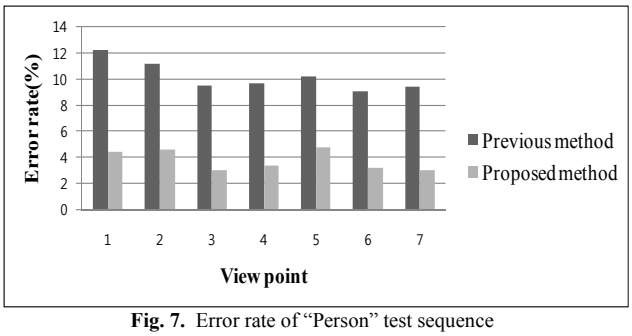
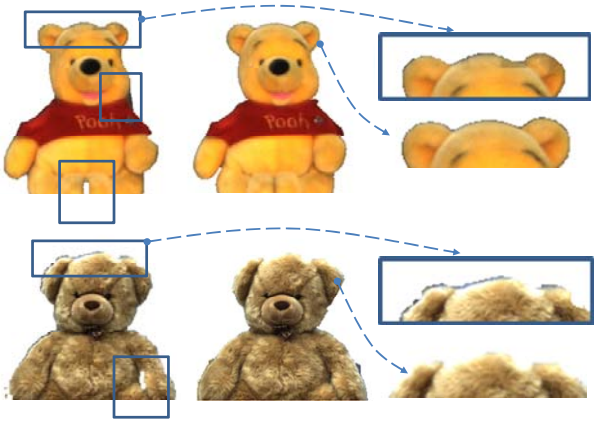


Fig. 7. Error rate of "Person" test sequence

As shown in Fig. 8, we compared extracted foregrounds by the previous and proposed methods. The previous method had problems on the foreground boundaries compared to the proposed method. While the proposed method extracted foreground boundaries more exactly than the previous method, the previous method was failed due to the color similarity between foreground and background at times.



(a) Previous method (b) Proposed method

Fig. 8. Comparison of segmented foreground objects

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we have proposed the multi-view image matting and compositing technique for the multi-view camera system. By using the multi-view camera system, we have reduced manual steps for the trimap generation. We have made the trimap from SAI and shared it for multi-view images to make alpha mattes by alpha estimation. Finally, we made alpha mattes by the edge labeling algorithm. Using the view-dependent alpha mattes, we extracted foreground objects of each view and composite them with corresponding multi-view backgrounds. In contrast to previous methods, the proposed method enables us to do the multi-view matting and compositing by considering foreground objects and their corresponding backgrounds. Finally, we have verified that multi-view composite images can generate 3-D scenes through the stereoscopic monitor. We plan to extend the narrow camera array to normal array by using the camera clustering technique for a general multi-view camera environment.

ACKNOWLEDGENENTS

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

REFERENCES

[1] Y.Y. Chuang, B. Curless, D.H. Salesin, and R. Szeliski, "A Bayesian Approach to Digital Matting," *Proc. of Computer Vision and Pattern Recognition*, pp. 264-271, 2001.

[2] T. Porter and T. Duff, "Compositing Digital Images," *Computer Graphics*, vol. 18, no. 3, pp. 253-259, 1984.

[3] A.R. Smith and J.F. Blinn, "Blue Screen Matting," *SIGGRAPH*, pp. 259–268, 1996.

[4] P. Hillman and J. Hannah, "Natural Image Matting," *Proc. of Vision, Video and Graphics*, pp. 213-216, 2005.

[5] M.A. Ruzon and C. Tomasi, "Alpha Estimation in Natural Images," *Proc. of Computer Vision and Pattern Recognition*, pp. 18-25, 2000.

[6] J. Sun, J. Jia, C.K. Tang, and H.Y. Shum, "Poisson Matting," *SIGGRAPH*, pp. 315-321, 2004.

[7] N. Joshi, W. Matusik, and S. Avidan, "Natural Video Matting Using Camera Arrays," *SIGGRAPH*, pp. 779–786, 2006.

[8] A. Isaksen, L. McMillan, and S.J. Gortler, "Dynamically Reparameterized Light Fields," *SIGGRAPH*, pp. 297–306, 2000.