IEICE Electronics Express, Vol.6, No.22, 1602–1607

Improving alpha matte with depth information

Ji-Ho Cho[1a)], Remo Ziegler[2], Markus Gross[3], and Kwan H. Lee[1]

[1] *Gwangju Institute of Science and Technology 1, Oryong-dong, Buk-gu, Gwangju 500–712, Korea*

[2] *LiberoVision AG, Technoparkstrasse 1 CH-8005 Zürich, Switzerland*

[3] *Haldeneggsteig 4/Weinbergstrasse ETH Zentrum CH-8092 Zürich, Switzerland*

a) *jhcho@gist.ac.kr*

Abstract: In this paper, we propose an automatic alpha matting method that improves the quality of alpha matte using a depth camera. Depth camera obtains both visual and depth information of moving objects in real time. Our matting method improves existing natural alpha matting methods by adding depth information. Experimental results demonstrate that our matting method generates better results than the previous approaches.

Keywords: alpha matting, depth camera, depth imaging

Classification: Science and engineering for electronics

References

[1] A. R. Smith and J. F. Blinn, "Blue screen matting," in *Proc. ACM SIGGRAPH 1996*, pp. 259–268, 1996.
[2] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. IEEE CVPR 2001*, pp. 264–271, 2001.
[3] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proc. IEEE CVPR 2006*, pp. 61–68, 2006.
[4] J. Wang and M. F. Cohen, "Optimized color sampling for robust matting," in *Proc. IEEE CVPR 2007*, pp. 1–8, 2007.
[5] M. Kawakita, K. Iizuka, T. Aida, T. Kurita, and H. Kikuchi, "Real-time three-dimensional video image composition by depth information," *IEICE Electron. Express*, vol. 1, no. 9, pp. 237–242, Aug. 2004.
[6] R. Gvili, A. Kaplan, E. Ofek, and G. Yahav, "Depth keying," in *Proc. SPIE Electronic Imaging Conference*, 2003.
[7] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 567–576, 2006.
[8] 3DV Systems, [Online] http://www.3dvsystems.com/
[9] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Flash matting," in *Proc. ACM SIGGRAPH 2006*, pp. 772–778, 2006.

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009

ELectronics EXpress

1 Introduction

The alpha matting is one of the most important tasks in photo and video editing, movie special effect, and digital contents production. This matting enables soft extraction of a foreground from a given image so that it allows a seamless composition of a foreground image with a new background image.

This problem is an inherently ill-posed since there are too many unknowns. For each pixel of the unknown region, there are seven unknowns (alpha and triplets for both foreground and background) and three equations (one per each channel). Therefore, a user-controlled background or user-interaction is required to solve matting problem. The bluescreen matting is a robust and widely used method that extracts the foreground object in front of the designated color screen by using color information [1]. It requires an expensive and well controlled studio environment with a blue or green color screen. As a result, actors cannot wear clothes of blue or green color.

In natural image matting, most recent methods [2, 3, 4] require a user-defined trimap or a scribble that segments an image into three regions: definite foreground, definite background and unknown. Given a trimap, all alpha values in the unknown region are solved from the known foreground and background information. Although most of them usually perform well, they generate an erroneous alpha matte when the foreground and background are visually similar.

Some researchers have developed a new approach using depth information called depth keying [5, 6]. It segments foreground objects using distance information without using a special background or a user-defined trimap. However, current depth keying systems require further alpha matting process to represent hairy objects.

In this paper, we propose a new system that improves matting result by adding depth information. Fig. 1 illustrates an overview of the proposed system. First, a depth camera captures both color and depth data. Then, a trimap is generated by dilation over the boundary of the depth image. Next, an alpha matte is estimated by color and depth information. Finally, a new

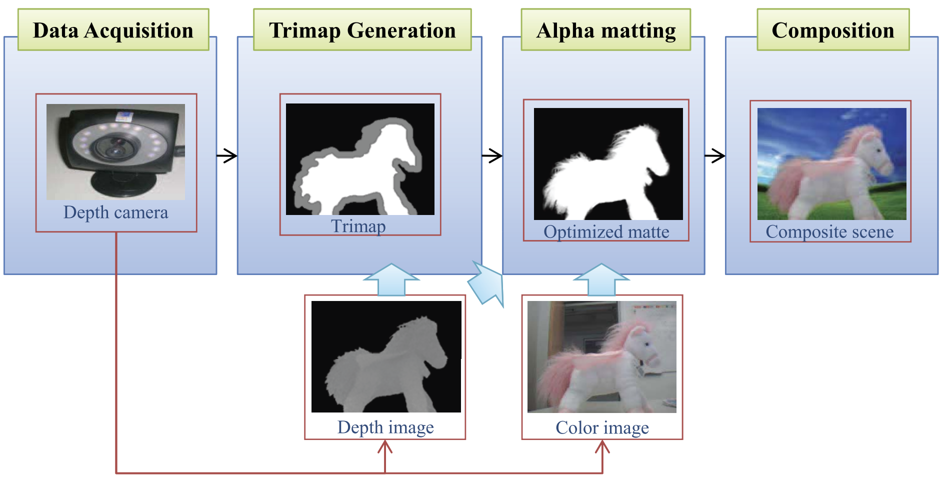


Fig. 1. Overall procedure

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009

IEICE Electronics Express, Vol.6, No.22, 1602–1607

scene can be synthesized by using the extracted foreground object.

2 The proposed method

2.1 Trimap generation

The depth map is used to automatically generate a trimap. All unknown areas belong to the region which is the border between the foreground object and the background. We obtain the unknown area T_U by applying the dilation operation on the boundary contour with the structuring element SE. The user should define the size and shape of SE for dilation operations considering the fuzziness of the foreground object. We use 15×15 circular matrix for SE and dilate the boundary region of the target depth images (320×240 resolution). Now that we have the color image, the depth image and the trimap for every frame. Therefore, we can automatically generate an alpha matte using the following matting algorithm.

2.2 Improving matte with depth information

Most color-based matting algorithms [2, 3, 4] work well for many natural scenes. However, all of them generate an erroneous alpha matte when the colors of the foreground and background are similar. In order to overcome this drawback, we use depth information that is captured independently of color similarity. We embed depth information into Robust matting [4] because Robust matting employs optimal sampling method and produces a good matte. We briefly summarize below how the alpha matte of a foreground object is extracted.

2.2.1 Initial alpha mattes generation

First, $n \times n$ sample pairs of the foreground and the background are collected along the boundaries of both known regions. Among them, high confidence sample pairs are chosen to obtain a more accurate and robust matte. The color-based alpha matte $\hat{\alpha}_c$ is computed by using the estimated foreground and the background:

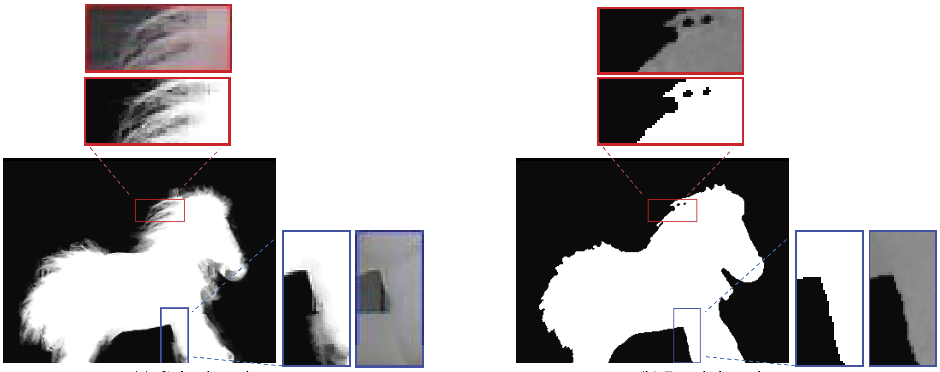
$$\hat{\alpha}_c = \frac{(I_c - \bar{B}_c) \cdot (\bar{F}_c - \bar{B}_c)}{\|\bar{F}_c - \bar{B}_c\|^2} \tag{1}$$

where I_c is the unknown pixel vector of the input color image. $\bar{F}_c$ and $\bar{B}_c$ are the sampled foreground and background vectors with the highest confidence values. More detail of the sampling method is described in [4]. Fig. 2 (a) shows the color-based matting result. The regions where foreground and background colors are visually different are accurate while the similar regions are poor. The poor regions are replaced by an alpha matte from the depth.

Depth information makes it possible to distinguish a foreground region from the background. However, it does not provide a fractional alpha because an alpha is not a linear combination of the depth of the foreground and the background. Therefore, depth-based alpha $\hat{\alpha}_d$ is obtained by converting a

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009

IEICE Electronics Express, Vol.6, No.22, 1602–1607



(a) Color-based matte (b) Depth-based matte

Fig. 2. Color and depth based matting result

gray-level depth map into a binary depth map. Fig. 2 (b) shows the depth-based matte. Note that the captured depth value is accurate even though its foreground and background are similar.

When the sampled background and foreground are visually different, the color-based matting produces the best matte. Therefore, we use color information in most cases and use depth information only when foreground and background are similar. The worst case is that the unknown pixel is hairy and its sampled colors are similar. We discuss this problem in section 3.

2.2.2 Combination of two mattes

We combine two alpha mattes (one is based on the color image and another is based on the depth image) using the color difference. Euclidean distance of two color vectors in the RGB space has been widely used to check the color difference. However, a short distance between two vectors in the RGB space does not necessarily mean that they are visually similar. In order to measure the color difference more accurately, we convert the RGB space into a perceptually uniform CIE Lab color space. The color difference ΔE between the foreground and the background is calculated as follows:

$$\Delta E = \sqrt{(L_f - L_b)^2 + (a_f - a_b)^2 + (b_f - b_b)^2} \tag{2}$$

Using this difference, we define an additional weight for color information in order to decide the region that we need to use the depth information:

$$w_c = \begin{cases} 0 & \text{if } \Delta E < T, \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

where T is the threshold of the color difference. We set this value between 6 and 10 in our system. Finally, the initial alpha matte is selectively combined using both color and depth based mattes using above weight:

$$\hat{\alpha} = w_c \hat{\alpha}_c + (1 - w_c)\hat{\alpha}_d \tag{4}$$

2.2.3 Matte optimization

We optimize the combined alpha matte to achieve local smoothness by graph-based optimization. To construct graph, we define two data weights $W_{i,F}$

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009

IEICE Electronics Express, Vol.6, No.22, 1602–1607

and $W_{i,B}$ between each node i and two virtual nodes F and B based on the combined alpha value:

$$W_{i,F} = \gamma \cdot \hat{\alpha}_i \tag{5}$$

and

$$W_{i,B} = \gamma \cdot (1 - \hat{\alpha}_i) \tag{6}$$

where γ is global smoothness factor. The more γ is low, the smoother a matte becomes. We set $\gamma = 0.1$ for all experiments.

For edge weight $W_{i,j}$ between two neighboring nodes i and j, the matting affinity function [3] is used. We add depth as the 4th channel with RGB channels. As a result, a matting affinity function becomes:

$$W_{i,j} = \sum_{k}^{(i,j)\in N_k} \frac{1}{|k|}\left(1 + (I_i - \mu_k)(\Sigma_k + \frac{\epsilon}{|k|} I_4)^{-1}(I_j - \mu_k)\right) \tag{7}$$

where N_k is the set of $m \times m$ neighboring windows containing node i and j, Σ_k is a covariance matrix, μ_k is a mean vector of r, g, b and d in each window N_k around k, and I_4 is a 4×4 identity matrix. We set $m = 3$ and $\epsilon = 0.00001$ in our experiment. Furthermore, we weight depth channel using $(1 - w_c)$ in order to use depth information only when foreground and background colors are similar.

Finally, the mattes are optimized by random walk optimizer [7]. Random walk is a graph-based optimization method that allows soft segmentation. More details of this graph construction and the optimization can be found in [4].

3 Results and discussion

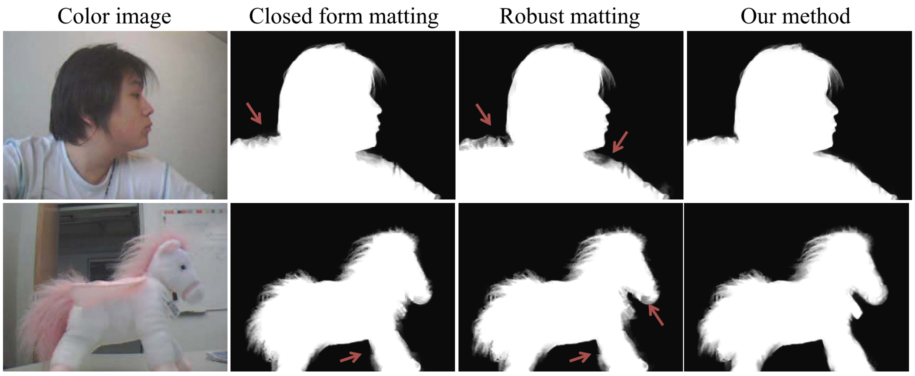
We tested the proposed method on three different hairy objects: a human model and two dolls. The images taken by a depth camera [8] have the resolution of 320 pixels in width and 240 pixels in height.

We compare the alpha matting results by our method with those by the previous color-based methods as shown in Fig. 3. The first column shows the RGB color images. Both color-based matting methods (closed form matting and Robust matting) generate erroneous results when the colors of the foreground and the background are similar as shown in the second and third columns in Fig. 3 (a). On the other hand, our method produces a better matte as shown in the last column. When the foreground and background are similar, closed form matting is better than Robust matting because robust matting is more sensitive to the sampled color pairs.

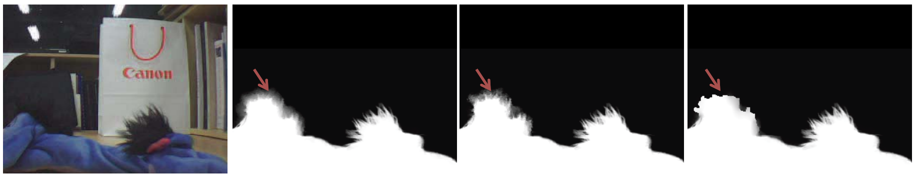
We have tested the worst case as shown in Fig. 3 (b). A donkey doll has the black fur on both its head and tail. In order to show the effect of color similarity of the foreground and the background on the matting algorithm, we set a black background on its head, and a white background on its tail. All matting methods generate pretty much the same and good results as shown in tail region. However, color-based matting methods generate a poor matte when the colors are similar as shown in the head area. In contrast,

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009

IEICE Electronics Express, Vol.6, No.22, 1602–1607



(a) Comparison with various matting methods



(b) Failure case

Fig. 3. Various alpha matting results

our method replaces this poor region with a depth value. Unfortunately, the captured depth map is inaccurate since this region is hairy. As a result, our method also produces not a good matte. Nevertheless, we believe its artifacts are less severe than the matte from color information. To pull a better matte in this case, further information such as flash/no-flash [9] image pair is required.

4 Conclusion

In this paper, we propose a new method that extracts a good matte even when the colors of the foreground and the background are similar by fusing the color and depth information. The results show that our matting method produces a better matte than the ones by the previous alpha matting methods. The extracted matte and depth can be use to generate a three-dimensional scene for a hairy object in the real world. We expect the proposed system to be used in various application areas such as 3DTV, game and movie industry.

Acknowledgments

This research was supported in part by the ITRC through RBRC at GIST (IITA-2009-(C1090-0902-0017)) and in part by the NRL (National Research Laboratory) funded by the Ministry of Education, Science and Technology of Korea (No. 20090083091).

DOI: 10.1587/elex.6.1602
Received October 08, 2009
Accepted October 27, 2009
Published November 25, 2009