Bayesian Image Matting Using Infrared and Color Cues

Layachi Bentabet and Hui Zhang

Computer Science Department, Bishop's University,
J1M 1Z7, Sherbrooke, Québec, Canada
{layachi.bentabet,hui.zhang}@ubishops.ca

Abstract. In this paper, we propose a new matting solution that combines the use of color and infrared cameras for matting applications involving human actors. The infrared camera facilitates the extraction of the initial trimap and provides additional information for the matte estimation. The approach proposed in this paper differs from the techniques proposed in the literature in many aspects. It employs thermal information for human actors, which proves to be useful and effective for matting when combined with color information. It also introduces a new technique for automatic trimap construction that is based on the temperature's difference between the foreground actor and the background objects. Finally, the matting step is carried out using a Bayesian approach which combines the color and the infrared inputs into a single criterion. The matting results accuracy shows that our approach is capable of tackling digital image and video matting problems.

Keywords: Image matting, trimap construction, cues combination, infrared imagery, background subtraction.

1 Introduction

The matting problem of separating a non-rectangular foreground image from a background image is a classical problem in image processing and analysis [1][2]. A common example is a film frame where an actor is extracted from the background to later be placed on a different background. When the original image is of high resolution and/or contains motion blur and grain, as what is usually used in visual effects industry, the matting becomes an under determined problem, for which a unique solution cannot be found. Images of this type are currently matted by help of user input, a process that is time consuming. To increase the quality of the mattes shot against arbitrary backgrounds, and also to reduce the amount of human interaction required to generate them, several matting techniques make use of additional imagery information.

In blue screen approaches [3], the matting problem is simplified by using a constant color background which normally is blue. To have a unique alpha solution for the blue screen matting problem, there should be no pure blue color in the foreground image. Blue Screen matting is most popular design of matting in TV studios & movies. In [4], a flash matting approach is proposed to extract alpha matte

A. Elmoataz et al. (Eds.): ICISP 2012, LNCS 7340, pp. 1–8, 2012.

by using flash/non-flash image pairs. This technique is based on the observation that the most noticeable difference between the flash and no-flash image is the foreground object, if the background scene is sufficiently distant. The algorithm is strongly based on the assumption that the foreground objects become brighter with the flash whereas the background objects remain the same. This assumption does not happen all the time in real scenes. Other possibilities which may cause the failure of the matting could be the low reflectance of the foreground surfaces and pixels' saturation. This approach also assumes that the input image pair is pixel-aligned. Thus, it will fail when the fine foreground structures have moved in the time interval between the two images. Other techniques, such as [5], propose the use of a camera system that builds the alpha matte using the parallax motion between the frames.

In this paper, we propose a new matting system that combines the use of infrared and color information. The combination of color and infrared cues has been studied recently for target tracking applications [6]. Many algorithms focusing on the thermal domain have been explored. These methods are based on the assumption that the objects of interest appear at a contrast from their surroundings in the scene [7][8]. Our solution is based on the assumption that a human body emits more heat than a background containing non living objects. We will show that the use of infrared allows automatic extraction of the trimap and accurate estimation of the alpha matte

2 The Matting Problem

2.1 Matting Equations

An input image I is composed of a foreground component F and a background component B as indicated in equation (1), where for the i^{th} pixel of the input image I_i is:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \,. \tag{1}$$

In the equation above, F_i is the foreground color, B_i is the background color and α_i is the pixel's foreground opacity. If $\alpha_i = 1$, the related pixel belongs to the foreground. If $\alpha_i = 0$, the related pixel belongs to the background. Otherwise we call it a mixed pixel. For a color image, C, equation (1) is generalized over RGB channels as follows:

$$C_R = \alpha_R F_R + (1 - \alpha_R) B_R \tag{2}$$

$$C_G = \alpha_G F_G + (1 - \alpha_G) B_G \tag{3}$$

$$C_B = \alpha_B F_B + (1 - \alpha_B) B_B \,. \tag{4}$$

In a color image, all quantities on the right-hand side of the above equations are unknown. We assume that alpha matte for red, green and blue channels are the same ($i.e. \alpha_R = \alpha_G = \alpha_B$). Therefore, for each pixel of a color image, there are three equations and seven unknowns. Thus matting is inherently an under-constrained problem.

2.2 Trimap Based Techniques

As mentioned before alpha matting is an ill-posed problem. Therefore, we need additional information about the image before to proceed with alpha estimation. Several approaches, such as Bayesian [9] and Robust matting [10], start by the user manually segmenting the input image into three regions, called trimap. A trimap is composed of three regions: a known foreground region, Ω_F, where $\alpha = 1$; a known background region, Ω_B, where $\alpha = 0$; and an unknown region, Ω_U, where $\alpha \in [0,1]$ (see figure 1). The foreground and background regions provide the additional information that is needed to estimate α in the unknown region.

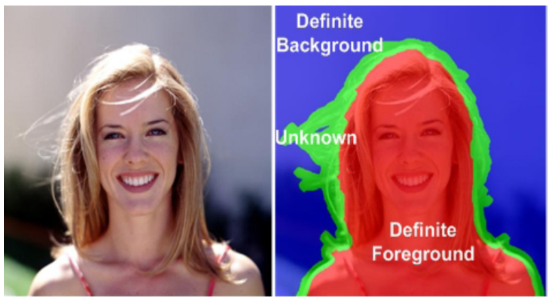

Fig. 1. Left image: original color image. Right image: user specified trimap.

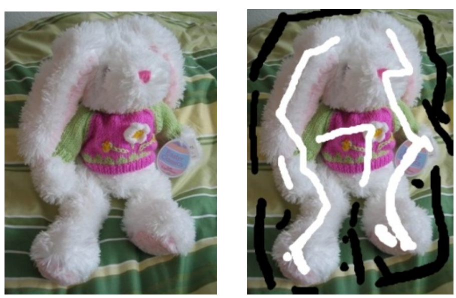

Fig. 2. Left image: input image. Right image: input image with scribbles constraints.

Recent techniques proposed the use of user-interface scribbles [11] as shown in figure 2. Instead of marking the whole image into three regions, the user puts some scribbles according to the eigenvectors of a Laplacian measure.

3 Bayesian Image Matting Using Infrared and Color Cues

3.1 Overview

In our approach, we propose to use an infrared camera as an additional source of information. Our research assumes that the foreground component contains only living subjects such as humans or animals. Consequently, the use of infrared sensor will help subtracting the background, and thus reduce the effect of similar colors in the background on the foreground matte. Instead of manually providing a trimap as most of the proposed methods in the literature, we use infrared images to automatically generate it. A general overview of our approach is presented in figure 3.

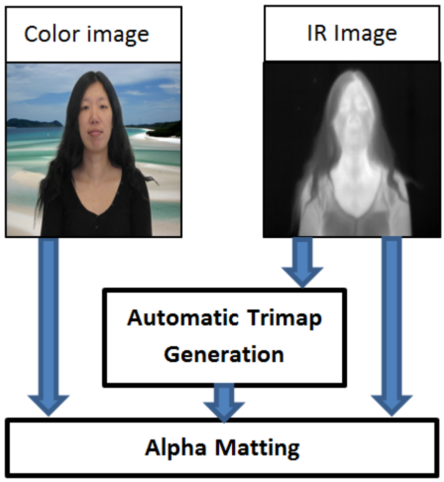

Fig. 3. Alpha matting using infrared and color sensors

The infrared and color images are assumed to be synchronized and spatially registered.

3.2 Automatic Trimap Generation

Our method starts with a foreground mask obtained by thresholding the entire infrared image. The threshold value is chosen according to the image histogram. As a result, the image domain is split into a foreground region and a region that contains both background and mixed pixels. In the next step, the unknown pixels are iteratively separated from the background.

The infrared level is characterized by the facts that: 1) it drops quickly at the border between foreground and background regions; 2) it decreases slowly for the pixels contained in the unknown region; and finally 3) it is low for the pixels in the background region. Given these considerations, our approach starts with a foreground mask that is iteratively propagated towards the background zone. At each iteration, new pixels are added to the unknown region if:

1) they are connected to the foreground or to the current unknown region, and,
2) their infrared levels are greater than a given threshold, and,
3) the difference between the infrared level of the new pixels and the infrared level of the pixels on the border of the unknown region is less than a given threshold.

Condition 1) ensures that the foreground and the unknown regions form together a single connected component. Condition 2) sets a threshold under which a pixel is considered as part of the background because of its low heat emission. Condition 3) reflects the smooth decrease of the infrared level in the unknown region. This corresponds to the human's temperature decrease as heat propagates in long thin parts such as the hair. The decrease rate is assumed to be lower than a given threshold. Therefore, if the transition at a given pixel is rapid, the pixel is considered as being part of the background. Our approach can be summarized as follows:

Let Ω_F^n and Ω_U^n be the foreground and unknown regions at iteration n. We define $\Omega_{F-U}^n = \Omega_F^n \cup \Omega_U^n$. The unknown region at iteration $n+1$ is given by:

$$\Omega_U^{n+1} = \Omega_U^n \cup \partial\Omega_{F-U}^n. \tag{5}$$

where $\partial\Omega_{F-U}^n$ is the set of pixels that are added to the unknown region. If pixel $x \in \partial\Omega_{F-U}^n$ then:

1) $x \notin \Omega_{F-U}^n$ and
2) $IR(x) > T_1$ and
3) $\exists z \in (\Omega_{F-U}^n \cap N_8(x))$ such that $||IR(z) - IR(x)| < T_2$

$N_8(\,)$ is the set of 8-Neighbors, T_1 and T_2 are thresholds set by the user.

3.3 Joint Bayesian Matting

In our algorithm, we consider the IR information as an additional channel similar to the RGB channels of the color image C. We thus form a 4D image $I_{4D} = (R, G, B, IR)^T$. Image matting is then written and solved for this image using a Bayesian framework.

In Bayesian estimation, we find the most likely estimates for F_{4D}, B_{4D} and α, given the observation IR and C. We can express this as a maximization of the posterior probability (MAP) $P(\alpha, F_{4D}, B_{4D}|I_{4D})$, and then use Bayes's rule with the log likelihood as follows:

$$\underset{\alpha, F_{4D}, B_{4D}}{arg\,max}\, P(\alpha, F_{4D}, B_{4D}|I_{4D}) \propto argmax\{L(I_{4D}|a, F_{4D}, B_{IR}) + L(F_{4D}) + L(B_{4D})\} \tag{6}$$

The log likelihood for alpha $L(\alpha)$ is assumed to be constant since we have no appropriate prior for α's distribution. The first two log likelihoods on the right hand side of (6) are used to measure the fitness of solved variables(α, F_{4D}, B_{4D}) with respect to matting equations. We model these terms by measuring the difference between the observed I_{4D} and the image that would be predicted by the estimated F_{4D}, B_{4D} and α.

$$L(I_{4D}|\alpha, F_{4D}, B_{4D}) = -\| I_{4D} - \alpha F_{4D} - (1-\alpha) B_{4D} \|^2 / \sigma_I^2$$

For $L(F_{4D})$ and $L(B_{4D})$, we follow the color sampling technique proposed in [9], where a group of nearby foreground and background pixels are collected to form an oriented Gaussian distribution. Thus:

$$L(F_{4D}) = -(F_{4D} - \overline{F_{4D}})^T \textstyle\sum_{F_{4D}}^{-1}(F_{4D} - \overline{F_{4D}})$$
$$L(B_{4D}) = -(B_{4D} - \overline{B_{4D}})^T \textstyle\sum_{B_{4D}}^{-1}(B_{4D} - \overline{B_{4D}})$$

$\overline{F_{4D}}$ and $\overline{B_{4D}}$ are the mean values of the foreground and background components. $\sum_{F_{4D}}^{-1}$ *and* $\sum_{B_{4D}}^{-1}$ are the covariance matrices. The minimization steps of equation (6) are detailed in [9] and can be summarized by the following algorithm:

Step 1: Fix α to solve for F_{4D} and B_{4D}:

$$\begin{bmatrix} \sum_{F_{4D}}^{-1} + I_{4D}\alpha^2/\sigma_{4D}^2 & I_{4D}\alpha(1-\alpha)/\sigma_{4D}^2 \\ I_{4D}\alpha(1-\alpha)/\sigma_{4D}^2 & \sum_{B_{4D}}^{-1} + I_{4D}(1-\alpha)^2/\sigma_{4D}^2 \end{bmatrix} \begin{bmatrix} F_{4D} \\ B_{4D} \end{bmatrix} = \begin{bmatrix} \sum_{F_{4D}}^{-1}\overline{F_{4D}} + I_{4D}\alpha/\sigma_{4D}^2 \\ \sum_{B_{4D}}^{-1}\overline{B_{4D}} + I_{4D}(1-\alpha)/\sigma_{4D}^2 \end{bmatrix}$$

Step 2: Fix F_{4D} and B_{4D} *to solve for* α:

$$\alpha = \frac{(I_{4D} - B_{4D}).(F_{4D} - B_{4D})}{\|F_{4D} - B_{4D}\|^2}$$

where I_{4D} is the 4×4 identity matrix. To maximize equation (6), we iteratively estimate α and $(F_{4D,}B_{4D})$ using steps 1 and 2 until changes between two successive iterations are negligible.

4 Experimental Results

The automatic trimap extraction method described in section 3.2 is applied on the infrared image of figure 4. The results are given in figure 5. The foreground is in white, the background in black and the unknown region in grey. These results show that our technique produces a single connected component for the three regions. The long hair details were successfully classified in the unknown region.

The joint Bayesian matting is applied on the unknown regions of the extracted trimaps and the results are presented in figure 5. The results demonstrate the ability of the developed technique to accurately estimate the alpha channel. As illustrated in figure 6, the level of detail for regions such as the hair is very accurate. It shows the transparency property of these areas, which is the main characteristic that allows a successful compositing of the actors on a new background.

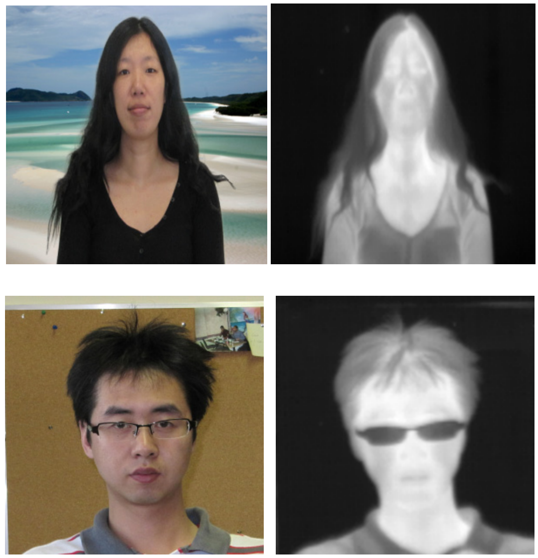

Fig. 4. Color and infrared image pairs

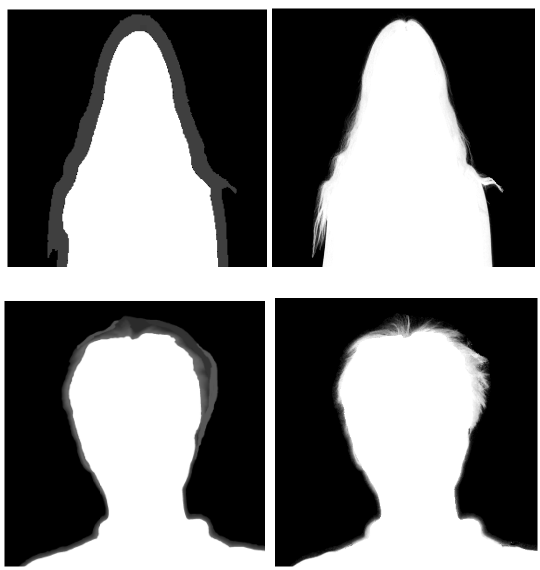

Fig. 5. Left: Automatic Trimap. Right: Alpha matte.

Fig. 6. Enlargement of the result in figure 4 that shows the details of the hair

5 Conclusion

In this paper, we have developed a Bayesian approache to solve the image matting problem. Though sharing a similar probabilistic view with [9], our approach differs in a number of key aspects. It uses MAP estimation to optimize color term and thermal term simultaneously. Also, the trimap is automatically generated from the thermal image. Our approach has an intuitive probabilistic motivation, is relatively easy to implement, and provide accurate matte results.

References

1. Fishkin, K., Barsky, B.: A family of new algorithms for soft filling. ACM SIGGRAPH 18(3), 235–244 (1984)
2. Porter, T., Duff, T.: Compositing digital images. ACM SIGGRAPH 18(3), 253–259 (1984)
3. Smith, A., Blinn, J.: Blue screen matting. ACM SIGGRAPH 30, 259–268 (1996)
4. Sun, J., Li, Y., Kang, S.-B., Shum, H.-Y.: Flash matting. ACM SIGGRAPH 25(3), 772–778 (2006)
5. Joshi, N., Matusik, W., Avidan, S.: Natural video matting using camera arrays. ACM SIGGRAPH 25(3), 779–786 (2006)
6. Airouche, M., Bentabet, L., Zelmat, M., Gao, G.: Pedestrian tracking using color, thermal and location cue measurements: a DSmT-based framework. Machine Vision and Applications (2011) (online first), doi:10.1007/s00138-011-0342-z
7. Yilmaz, A., Shafique, K., Shah, M.: Target tracking in airborne forward looking infrared imagery. Image and Vision Computing 21(7), 623–635 (2003)
8. Fernández-Caballero, A., Carlos-Castillo, J., Serrano-Cuerda, J., Maldonado-Bascón, S.: Real-time human segmentation in infrared videos. Expert Systems with Applications 38, 2577–2584 (2011)
9. Chuang, Y.-Y., Curless, B., Salesin, D.-H., Szeliski, R.: A bayesian approach to digital matting. IEEE CVPR 2, 264–271 (2001)
10. Wang, J., Cohen, M.: Optimized color sampling for robust matting. IEEE CVPR, 1–8 (2007)
11. Levin, A., Lischinski, D., Weiss, Y.: A closed form solution to natural image matting. IEEE TPAMI 30, 228–242 (2008)