Image Matting Based on Mutual Information

Xiaozhou Zhou and Pierre Boulanger
Department of Computing Science
University of Alberta, Edmonton
Alberta, Canada, T6G 2E8
{xzhou3, pierreb} @ ualberta.ca

***Abstract*- In this paper, we propose a novel framework to solve the image matting problem. We design a temporary image based on the estimated foreground and background colors for unknown pixels as well as an initial matte. The similarity of the temporary image and original image is modeled as an energy function in the Markov Random Field (MRF). The global optimized matte is obtained by minimizing the energy function. Therefore, image matting is converted to how to maximize the similarity of the original image and the temporary image. The experiments demonstrate that the proposed method could produce high quality mattes and it is also more effective compared to other top ranking methods.**

I. INTRODUCTION

Image matting is a process of extracting foregrounds from images. The extracted foreground is then pasted on another background. This is one of the key steps in image editing and special effects in film industry. Image matting seems similar with image segmentation. However, in image segmentation, each pixel belongs to either the foreground or the background; while image matting thinks each pixel is a combination of foreground color and background color, which makes the boundary of foreground smoother and retains tiny details, such as hairs and semi-transparent stuff. Image matting could be expressed mathematically as [1]:

$$I(i,j) = \alpha_{i,j} * F(i,j) + (1 - \alpha_{i,j}) * B(i,j), \qquad (1)$$

where (i,j) is the 2D image coordinates of a pixel and $I(i,j)$represents the color of the pixel (i,j). Similarly, $F(i,j)$ and $B(i,j)$ are the foreground and background color of pixel (i,j) respectively. The parameter $\alpha \in [0,1]$ is a weighing value for each pixel between the foreground and the background. It measures how many percent of the color is from the foreground and how many percent of the color is from the background. $\alpha_{i,j} = 1$ means that the color of pixel is totally from the foreground, while $\alpha_{i,j} = 0$ means that the pixel is totally from the background. Image matting usually goes with a trimap which is a three-color image. The white means definite foreground (DF), the black represents definite background (DB) and gray is the unknown region, as listed in Figure 2 (b). The result of image matting is a gray scale image, called "matte" in which the value of each pixel is α, scaled from [0,1] to [0, 255].

Mutual Information [2] was first coined by Shannon to measure the dependence of two random variables. It has been widely applied to Image Processing and Computer Vision area to compare the similarity of two images. We found that with estimated $F'(i,j)$ and $B'(i,j)$, image matting problem could be converted to another one: how to find the α values in order to reach the highest similarity between images I and I':

$$I'(i,j) = \alpha'_{i,j} * F'(i,j) + (1 - \alpha'_{i,j}) * B'(i,j). \qquad (2)$$

Therefore, in this paper, we propose a novel framework to solve image matting by maximizing the similarity of two images. I' is the temporary image constructed by estimating the foreground and background colors for each pixel in the unknown region. Then, the similarity between I and I' is measured by Mutual Information formulated as en energy function in MRF. Finally, the energy function solved by Graph Cut to reach the global optimization.

The rest of this paper is organized as: Section 2 introduces the related literature. Details of the proposed method are in Section 3. We will explain what Mutual Information is and how Mutual Information is re-formatted in MRF and solved by Graph Cut. The efficiency and robust are demonstrated in Section 4. Section 5 gives the conclusion and future work.

II. RELATED WORK

In the early stage, image matting was applied to extract objects with known backgrounds, such as Blue Screen Matting [3]. It photographed the same object with two known backgrounds and solved two matting functions with two unknown parameters α and F. Unfortunately, the requirement for known backgrounds limits this method in studio environment.

Later on, the research was focus on the matting of natural images, which had more complex texture patterns and unknown backgrounds. Poisson matting [4] assumed the foreground and background were smooth in a narrow band, thus the gradients of the colors in these regions were zeros.

The matting function then could be solved as a Poisson equation. Bayesian Matting [5] modeled the foreground and background colors by Bayesian framework and used Maximum A Posterior (MAP) to solve it. Trimaps used in this kind of methods are usually made manually which is time-consuming, especially in case the foreground with a complex shape or texture, such as mesh and smoke.

Robust Matting [6] suggested more robust color estimation by assigning a confidence value to each color sample and only those samples with high confidence values were selected to form the energy function.

Lazy Snapping [7] and GrabCut [8] worked on how to get trimaps automatically. Both of them used Graph Cut to segment the foreground and eroded/dilated the boundary of segmented foreground to create the unknown region. The user inputs were just a few strokes or a bounding box. However, hard segmentation always ignored fine features, which made it hard to guarantee the quality of trimap.

All trimaps based matting methods suffer from the same problem: the trimap must be as fine as possible to get a good matte. From above discussion, the making of trimaps is either tedious or unreliable, thus most recent direction switched to research how to get rid of the trimap. Easy Matting [9], Closed-Form Matting [10] and Iterative Belief Propagation (BP) [11] all needed only a few scribbles to indicate the foreground and background.

[9] proposed an iteratively framework. It was called "easy" because even for video matting, it also only asked a few strokes for some key frames. In each iteration, the unknown region was modeled as MRF which could be solved by energy minimization. Results were then used to update the estimation of color samples. The next iteration started based on the updated color samples. [11] also used an iterative frame and it divided pixel into U_c and U_n to represent pixels been processed or not. In each iteration, pixels in U_n but close to ones in U_c were added to U_c. Then, a model of MRF was built on pixels in U_c and solved by Belief Propagation. Iterations repeated until no pixel left in U_n or convergence. [10] supposed the foreground and background colors could be fit in a linear model in local area so that turned the matting problem to solve a quadratic cost function.

III. PROPOSED ALGORITHM

The proposed method consists of 3 parts: 1) Estimate the foreground and background colors for pixels in unknown region and form the temporary image by an initial matte. 2) Compute the Mutual Information of original image and temporary image, and construct the energy function. 3) Solve the energy function by Graph Cut. In this section, first, we introduce what is Mutual Information. Then, explain how to estimate the foreground and background colors and how to use Mutual Information in our framework.

A. Mutual Information

The entropy is one of the fundamental concepts in Information Theory. It measures the uncertainty of variables. Lower entropy means the variable is more predictable. The entropy of a random discrete variable X is defined by:

$$H(X) = -\sum_{x \in X} p(x) log p(x) \ . \qquad (3)$$

$p(x)$ is the probability distribution of x. For two random discrete variables X and Y. The joint entropy is defined by:

$$H(X,Y) = -\sum_{y \in Y} \sum_{x \in X} p(x,y) log p(x,y) \ . \qquad (4)$$

The joint entropy $H(X,Y)$is a statistics that summarizes the degree of dependence of X on Y.

Mutual information uses information entropy to measure relationship between two variables. The higher the Mutual Information is, the more information about X could be gained by knowing Y. Mutual Information is calculated by:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y). \qquad (5)$$

$MI(X,Y)$ reaches the maximum when X equals to Y.

B. Estimation of Foreground and Background

There are many ways to estimate the foreground and background colors for unknown pixels. Since in this paper, the importance is to explain the framework how to solve the image matting problem by maximizing the similarity of images. Therefore, with respect to the foreground and background estimation, we just employ the simplest one, which was introduced in Poisson Matting [4]. The estimated foreground color of an unknown pixel is the color of nearest pixel belonging to the definitely foreground region; while the estimated background color of an unknown pixel is the color of nearest pixel belonging to the definitely background region. This simplest estimation works well where the foreground and background colors are smooth around the unknown pixel. Nonetheless, it is not robust to noise. Statistic models based on color samples are more reliable. Moreover, the estimation of foreground and background could also be integrated into the energy function. In this way, we could also extend our method to scribble based methods and no need to estimate colors in advance.

The matte is initialized as 255 in DF, 0 in DB, and 0 in the unknown region. After this step, the temporary image I' is constructed by (2).

C. Energy Minimization of Mutual Information

In the proposed framework, we model the pixels as a Markov Random Field. (For trimap based methods, we only need to model pixels in the unknown region.) The energy function is defined as:

$$E(p) = E_{data}(p) + \lambda E_{smooth}(p) \ . \qquad (6)$$

Next, we show how to formulate $MI(X,Y)$ into the data term $E_{data}(p)$ based on original image I and temporary image I'. According to [12], $H(I)$ and $H(I')$ could be regarded as constant. Only the joint entropy needs to be concerned in data term which is:

$$E_{data}(\alpha) = \sum_p D_p(\alpha_p) = -MI(I,I') \approx H(I,I'). \qquad (7)$$

Mutual information reaches the maximum when two images are totally the same. Thus, a negative sign is added in front of Mutual Information to maximize the Mutual Information by minimizing the energy function. $D_p(\alpha_p)$ is the measurement of dissimilarity of two pixels $I(p)$ and $I'(p,\alpha)$. $I(p)$ is the intensity value of pixel p in image I. $I'(p,\alpha)$ is the pixel value of p in I' according to (2).

The probability distribution is defined as:

$$p\big(I(p), I'(p,\alpha)\big) = \left[\frac{1}{|P|}\sum_p T[(c_1,c_2) = (I(p), I'(p,\alpha))]\right] \otimes G(I(p), I'(p,\alpha)). \qquad (8)$$

$\sum_p T[(c_1,c_2) = (I(p), I'(p,\alpha))]$ is the way to construct the histogram of two images. c_1 and c_2 are the colors in [0,255]. $T[\cdot]$ is 1 if the argument is true, otherwise, $T[\cdot]$ is 0; thus the bin (c_1,c_2) plus one if $I(p) = c_1$ and $I'(p,\alpha) = c_2$. G is 2D Gaussian function. $|P|$ is the number of pixels in an image. Please refer to [12] for more details. The data term is:

$$E_{data}(\alpha) = -\sum_p\left(\frac{1}{|P|}\log p(I(p), I'(p,\alpha))\right) \otimes G(I(p), I'(p,\alpha)). (9)$$

The smooth term defined as following to keep the continuity in neighborhood,

$$E_{smooth}(p) = \sum_{q\in N(p)}(\alpha(p) - \alpha(q))^2. \qquad (10)$$

We use α- expansion Graph Cut to solve the energy function. The α is set as 100 levels which corresponds to the matte value from 0.0 to 1.0 with interval 0.01. In our experiment, λ is set as 0.2.

IV. RESULTS AND DISCUSSION

In this section, we first compare our method with several best matting algorithms in Figure 1: Random Walk, Poisson, Closed-Form Matting, Knockout2, Interactive BP, Bayesian Matting and Robust Matting. Those results are from [6]. Our results still rank high among those top methods. Then, we use more natural images to test our method and the results are listed in Figure 2. In the experiments, we use trimaps to indicate the unknown regions and the simplest way for estimation (details in 3.2), but the results are impressive already. It demonstrates that this framework is a reasonable way to solve the matting problem.

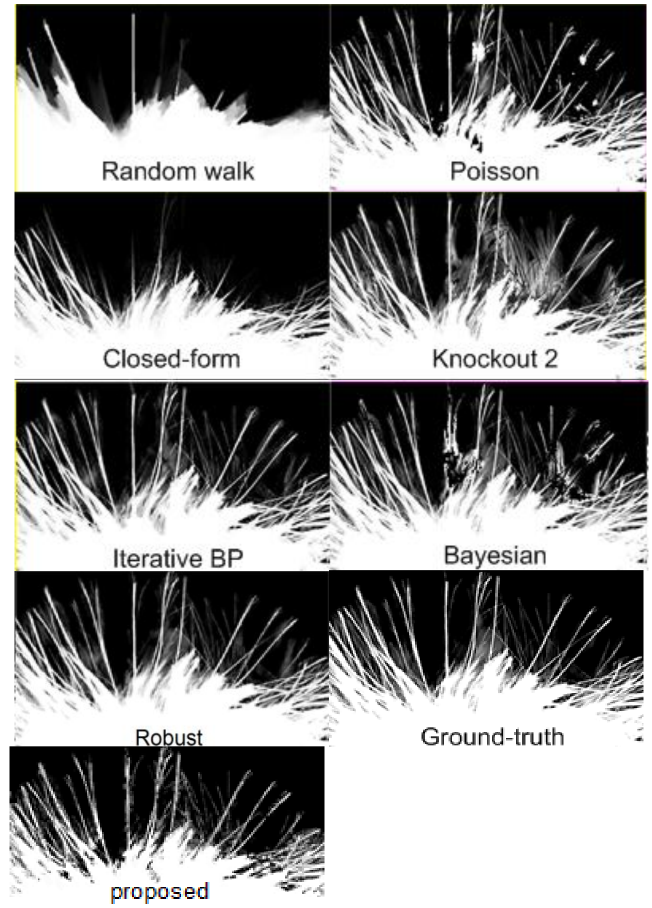

Figure 1. Comparison of mattes of proposed method with other algorithms

We could notice that the results rely heavily on the estimation of foreground and background colors. If we utilize more precise model to assign and update the foreground and background colors, we are pretty sure that the results will be even better. The proposed method could be also extended to an iterative frame. The estimation and matte are updated after each iteration.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel idea to solve the matting problem by maximizing the similarity of two images: original image and temporary one with estimated foreground and background colors. The similarity is modeled by Mutual Information and the energy function is minimized by Graph Cut. The experiments demonstrate the proposed method is effective in all the test images. And it is also one of the best matting algorithms published so far.

In the future, we plan to design a statistic based model to estimate the foreground and background colors from color samples, which makes the estimation more powerful.

(a) Original Test Images

(b) Trimaps

(c) Results of Proposed Method (Matte)

(d) Ground Truth

Figure 2. Results of proposed method

REFERENCES

[1] J. Wang and M. Cohen, “Image and video matting: a survey”. Foundations and Trends in Computer Graphics and Vision, vol.3, pp.97-175, Janary, 2007.

[2] C.E. Shannon, “A mathematical theory of communication”. Bell Systmes technical Journal, 27, pp. 379-423, 623-656, 1948.

[3] A.R. Smith and J.F. Blinn. “Blue Screen Matting”. In Proceedings of ACM SIGGRAPH, pp. 259-268, 1996.

[4] J. Sun, J.Y. Jia, C.K. Tang and H.Y. Shum. “Poisson Matting”. In Proceedings of ACM SIGGRAPH, pp. 315-321, 2004.

[5] Y.Y. Chuang, B. Curless, D. Salesin and R. Szeliski. ‘A Bayesian Approach to Digital Matting”. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 264-271, 2001.

[6] J. Wang and M. Cohen, “Optimized Color Sampling for Robust Matting”. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[7] Y. Li, J. Sun, C.K. Tang and H.Y. Shum, “Lazy snapping”. In Proceedings of ACM SIGGRAPH, pp. 303-308, 2004.

[8] C. Rother, V. Kolmogorov and A. Blake, “Grabcut – interactive foreground extraction using iterated graph cut”. In Proceedings of ACM SIGGRAPH, pages 309–314, 2004.

[9] Y. Guan, W. Chen, X. Liang, Z. Ding and Q. Peng. “Easy Matting: A Stroke Based Approach for Continuous Image Matting”. In Proceedings of Eurographics, pp. 567-576, 2006.

[10] A. Levin, D. Lischinski and Y. Weiss, “A closed form solution to natural image matting”. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 61-68, 2006.

[11] J. Wang and M. Cohen, “An iterative optimization approach for unified image segmentation and matting”. In Proceedings of International Conference on Computer Vision, pp. 936–943, 2005.

[12] J. Kim, V. Kolmogorov and R. Zabih, “Visual Correspondence using Energy Minimization and Mutual Information”. In Proceedings of International Conference on Computer Vision, vol.2, pp. 1003-1010, 2003.