IMAGE MATTING IN THE FRAMEWORK OF QUANTIFICATION IV

Takumi Kobayashi *Tadaaki Hosaka* *Nobuyuki Otsu*

National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Japan

ABSTRACT

Image matting and segmentation, which are used to extract a foreground object from the background, are primary techniques for digital image and video editing. In digital matting, the transparency of the foreground object is considered, while segmentation performs a rigid extraction of the object. Recently, several algorithms for matting and segmentation problems have been proposed and have provided high-quality results. In this paper, we propose a unified formulation for image matting in the framework of the method of Quantification IV based on a review of the previous studies from this framework. Our method also utilizes discriminative information provided by a user (a few strokes drawn by the user). The experimental results show a favorable matting performance.

Index Terms— Matting, Object Extraction, Image Editing, Quantification IV, SVM classifier

1. INTRODUCTION

As a tool for image and video editing, image matting is used to extract a foreground object from the background and place it into a new image in such a way that it appears natural. The matting problem is to estimate the opacity (called the *alpha* value) and foreground and background elements at each pixel, which are related each other by the following equation:

$$\boldsymbol{C}_i = \alpha_i \boldsymbol{F}_i + (1 - \alpha_i)\boldsymbol{B}_i, \tag{1}$$

where $\alpha_i \in [0, 1]$ represents the opacity; $\boldsymbol{C}_i$, the color vector in the image. $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$ represent the foreground and background color vectors at pixel i, respectively. The matting problem for natural images is inherently ill-posed since there are three observations (R,G,B in $\boldsymbol{C}_i$) and seven unknowns to be estimated in Eq.(1). Several algorithms have been proposed in the computer vision community to deal with the ill-posedness, and they have shown some high-quality results. In these algorithms, some user interactions are required for indicating the foreground object that is to be extracted; they also function as clues (constraints) for solving the problem.

As shown in Fig.1, there are two types of user interactions: *trimap* [1][2] and *strokes* [3][4]. The degree of user interaction in the strokes type is much less than that in the trimap type. The region of alpha estimation, however, is larger in the strokes type, which makes the matting problem more difficult. In [3], the alpha estimation is iteratively propagated from the strokes by using belief propagation. Levin *et al.* [4] have assumed that in a local window, $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$ lie on a straight line in RGB color space, respectively, and have transformed the above ill-posed problem into a closed-form expression by using a least square solution for $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$. Although Grady *et al.* [2] have used a trimap, their formulation based on the random walker algorithm is also valid for strokes type interactions. Moreover, the resulting formulation is almost the same as that in [4].

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

Fig. 1. (a) Original image. (b) Trimap: white/black pixels are the fore/background, respectively, and the alpha values in the gray pixels are estimated. (c) Strokes: a user draws red/blue strokes in fore/background region, respectively. (d) Example of an alpha matte. The opacities around the boundary are represented as gray-level alpha values.

Segmentation is a special case of image matting where the alpha values in Eq.(1) are limited to either 0 (background) or 1 (foreground). Some of the recent studies on segmentation are based on Graph Cuts [5][6]. The user interactions in these are of the strokes type. On the other hand, the Normalized Cut [7] has been proposed for region segmentation, which does not extract a foreground object but segments the regions in an image without the aid of user interactions.

In this paper, we first review the previous studies on both matting and segmentation problems in terms of the framework of Quantification IV (Q-IV) [8]. Based on this review, we propose a unified formulation for image matting, which naturally incorporates several constraints in the framework of Q-IV. One of the constraints includes discriminative information, which enables the unified formulation to deal with *difficult* images. The global solution of this formulation is also obtained.

Method	Objective	Constraint
Q-IV [8]	$\min \sum_{ij} s_{ij}(\alpha_i - \alpha_j)^2$	(I) $\sum_i \alpha_i^2 = 1$
Lazy Snapping [5]	$\min \sum_{ij} s_{ij}^L(\alpha_i - \alpha_j)^2$	$\alpha_i \in \{0,1\}$, (II) $\min \sum_i \{g_i^b \alpha_i + g_i^f(1-\alpha_i)\}$, (III) $\alpha_k = \begin{cases} 1 & (k \in \mathcal{P}_f) \\ 0 & (k \in \mathcal{P}_b) \end{cases}$
Normalized Cut [7]	$\min \sum_{ij} s_{ij}^N(\alpha_i - \alpha_j)^2$	(I') $\sum_i d_i^N \alpha_i^2 = 1$
Closed Form [4]	$\min \sum_{ij} s_{ij}^C(\alpha_i - \alpha_j)^2$	(III) $\alpha_k = \begin{cases} 1 & (k \in \mathcal{P}_f) \\ 0 & (k \in \mathcal{P}_b) \end{cases}$

Table 1. Comparison of formulations in the framework of Q-IV

2. FRAMEWORK OF QUANTIFICATION IV

We first describe the formulation of the method of Q-IV [8]; then, we review some previous studies in terms of Q-IV.

2.1. Quantification IV

In the method of Q-IV, samples are scaled appropriately according to the similarities among them, which is related to multidimensional scaling [9].

For n samples $\{O_1, \cdots, O_n\}$, we assume that a similarity s_{ij} is provided between O_i and O_j. Then, the problem is to obtain the coordinate α_i onto which sample O_i is mapped according to the similarity s_{ij}. This is formulated as

$$\min_{\boldsymbol{\alpha}} \sum_{i,j} s_{ij}(\alpha_i - \alpha_j)^2, \quad \text{s.t.} \quad \sum_i \alpha_i^2 = 1. \tag{2}$$

By using a Lagrange multiplier η, this reduces to an eigenvalue problem:

$$(\boldsymbol{D} - \boldsymbol{S})\boldsymbol{\alpha} = \eta \boldsymbol{\alpha}, \tag{3}$$

where $\boldsymbol{S}$ and $\boldsymbol{D}$ represent a symmetric similarity matrix ($\boldsymbol{S} = \{s_{ij}\}$) and a diagonal matrix ($\boldsymbol{D} = \text{diag}\{d_i = \sum_j s_{ij}\}$), respectively. The appropriate scaling for the samples consists of the second (or latter) smallest eigenvectors in Eq.(3). The minimization problem (2) implies that two samples with high similarity are positioned close to each other and are subject to the constraint of unit variance, which reduces the ambiguity of the scale of the coordinates. The remaining ambiguity of translation is eventually reduced because the second or latter smallest eigenvectors are orthogonal to $\mathbf{1}$ (vector), i.e., the mean of samples is 0.

The formulation itself corresponds to that of a spectral clustering which is based on a cut of graph. However, the purpose of Q-IV is to scale samples, and not to cluster them; this facilitates the interpretation of the roles of the constraints and continuities of the alpha values in image matting.

2.2. Review of Previous Studies

We review some previous studies on segmentation and matting problems in terms of the framework of Q-IV. Here, the sample O_i represents a pixel in an image, and the notations $\mathcal{P}_f$ and $\mathcal{P}_b$ represent the sets of pixels marked by a user with strokes as being definitely foreground and background, respectively. We describe these formulations briefly and summarize them in Table 1.

Segmentation The coordinate values $\boldsymbol{\alpha}$ are regarded as fore/background labels: α=1 for the foreground and α=0 for the background. The formulation used in *Lazy Snapping* [5] is similar to Eq.(2), and constraint (II) indicates that the energy defined at each pixel is possibly minimized. Energy (II) is based on the affinities to the fore/background, which is minimized if alpha obeys the affinity, and the similarity s_{ij}^L is derived from the smoothness term in [5]. The formulation of [6] is similar to that of [5]. In the *Normalized Cut* [7], a reverse transformation from Eq.(3) to Eq.(2) can be performed, and constraint (I') is regarded as *weighted* (I) in Q-IV.

Matting We focus on the estimation of alpha values $\boldsymbol{\alpha}$ as in the previous studies [2][4]. The coordinate of α_i may be easily interpreted as the probability of being part of the foreground at each pixel [2]. Constraint (III) determines the approximate boundaries of the alpha values ($\alpha_i \in [0,1]$). The formulation of the *Closed Form* [4] is similar to Eq.(3) and can be reversely transformed to Eq.(2) with constraint (III). The similarity s_{ij}^C used in [4] is

$$s_{ij}^C = \sum_{k|(i,j)\in w_k} \frac{1}{|w_k|}\Big\{1 + (\boldsymbol{C}_i - \boldsymbol{\mu}_k)(\boldsymbol{\Sigma}_k + \frac{\epsilon}{|w_k|}\boldsymbol{I})^{-1}(\boldsymbol{C}_j - \boldsymbol{\mu}_k)\Big\}, \tag{4}$$

where w_k denotes k-th local window, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the mean vector and the covariance matrix of color vectors in the local window, respectively, and ϵ represents a regularization parameter.

The framework of Q-IV is interpreted as follows: the objective function, $\min \sum s_{ij}(\alpha_i - \alpha_j)^2$, forms a system of $\boldsymbol{\alpha}$ values by taking into account the similarities; then, several constraints actually define the coordinates (scaling) such as $\alpha = 1/0$ for indicating definite fore/background. These constraints are classified into three types: *hard*, *soft* and *mild* constraints. The hard constraint corresponds to the user-drawn strokes (III). This specifies the actual value (1 or 0) for the pixels that have been selected by the user through interactions. The soft constraints are those of Q-IV and the Normalized Cut (I, I'), which limit the distribution of $\boldsymbol{\alpha}$ according to the (weighted) variance. The mild constraint is that of Lazy Snapping (II). This constraint is an intermediate between the

hard and soft constraints and possibly places α close to either 1 or 0 according to the affinities in a manner similar to that of regularization terms.

3. PROPOSED METHOD FOR IMAGE MATTING

Based on the above review, we propose a unified formulation for the matting problem in the Q-IV framework.

3.1. Formulation

The framework of Q-IV is particularly suitable for the matting problem due to the continuity of the alpha values. As seen in the review, several constraints are naturally incorporated in the framework. Our formulation for estimating alpha values is as follows:

$$\min_{\boldsymbol{\alpha}} \sum_{ij} s_{ij}^C (\alpha_i - \alpha_j)^2 \tag{5}$$

$$\text{s.t.} \quad \min \sum_{i \in \mathcal{P}} d_i \{\Omega_i^b \alpha_i^2 + \Omega_i^f (1-\alpha_i)^2\} \tag{6}$$

$$\text{and} \quad \alpha_k = \begin{cases} 1 & (k \in \mathcal{P}_f) \\ 0 & (k \in \mathcal{P}_b) \end{cases}, \tag{7}$$

where Ω_i^f and Ω_i^b indicate the affinities to the foreground and background at pixel i, respectively. This formulation includes both the *hard* and *mild* constraints. As described in the previous section, only the hard constraint (7) has been employed for the matting problem and it is a natural extension to add more constraints (here, the mild constraint (6)) in terms of Q-IV. The minimization of Eq.(6) is possibly performed as in Lazy Snapping. The similarity s_{ij}^C of Closed Form [4] is simply employed, and we assume that $s_{ii}^C = 0$ which does not affect the final solution at all.

The affinities are defined as follows similarly to [10],

$$\Omega_i^f = \frac{I[f(\boldsymbol{g}_i)]}{\max_{j \in \mathcal{P}_f} f(\boldsymbol{g}_j)}, \quad \Omega_i^b = \frac{I[-f(\boldsymbol{g}_i)]}{\max_{j \in \mathcal{P}_b} -f(\boldsymbol{g}_j)}, \tag{8}$$

where $f(\boldsymbol{g}_i)$ represents the output of the SVM classifier with a Gaussian kernel $\exp(-\gamma||\boldsymbol{g} - \boldsymbol{g'}||^2)$ for the vector $\boldsymbol{g}_i$, and $I[x]$ is equal to x for $x > 0$ and 0 otherwise. The vector $\boldsymbol{g}_i$ at pixel i is defined as the concatenation of color vectors of pixel i and the 8-neighboring pixels, or the color vector of pixel i. It is characterized by the number of concatenated pixels ($N \in \{1, 9\}$) which is a parameter decided by a user. From the definition (8), the affinities range from 0 to nearly 1 and are based on discriminative information by SVM classification. It is noted that the discriminative information for fore/background is naturally introduced into our matting formulation through the mild constraint, whereas previous methods on image matting rarely include it. Thus, our method deals with some *difficult* images for which the other methods have failed. Furthermore, the global solution is also obtained in spite of the addition of the constraints as described below.

The mild constraint (6) may be regarded as the combination of the constraints of Lazy Snapping and Normalized Cut; at each pixel, the alpha value is constrained to be closer to either 1 or 0 according to the affinities Ω^f and Ω^b, and the constraint is weighted by d_i. As described below, the *weighted* mild constraint also works on computation. It is also noted that this is a quadratic form of α. It encourages the alpha values to take intermediate values $0 < \alpha < 1$, because the alpha values loose their sparseness due to the second order, whereas the first order would enforce sparseness, nearly resulting in segmentation ($\alpha = 0/1$).

3.2. Computation

With a (balancing) parameter $\lambda \geq 0$, the proposed formulation is transformed to the following:

$$\min_{\boldsymbol{\alpha}} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix}^T \boldsymbol{L} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix} + \lambda \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix}^T \boldsymbol{D}(\boldsymbol{\Omega}^f + \boldsymbol{\Omega}^b) \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix} - 2\lambda \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix}^T \boldsymbol{D}\boldsymbol{\Omega}^f \mathbf{1} + \lambda \mathbf{1}^T \boldsymbol{D}\boldsymbol{\Omega}^f \mathbf{1}, \tag{9}$$

where $\boldsymbol{\Omega}^f = \text{diag}(\Omega_1^f, \cdots, \Omega_n^f), \boldsymbol{\Omega}^b = \text{diag}(\Omega_1^b, \cdots, \Omega_n^b)$, and $\boldsymbol{\delta}$ is a vector composed of the alpha values satisfying the hard constraint (7). Since this is the quadratic form of α, the global minimum solution is obtained by linear equations:

$$\{\boldsymbol{L_0} + \lambda \boldsymbol{D_0}(\boldsymbol{\Omega_0^f} + \boldsymbol{\Omega_0^b})\}\boldsymbol{\alpha} = \lambda \boldsymbol{D_0}\boldsymbol{\Omega_0^f}\mathbf{1} - \boldsymbol{L_1}\boldsymbol{\delta}, \tag{10}$$

where

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L_0} & \boldsymbol{L_1} \\ \boldsymbol{L_1}^T & \boldsymbol{L_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D_0} - \boldsymbol{S_0} & \boldsymbol{D_1} - \boldsymbol{S_1} \\ \boldsymbol{D_1}^T - \boldsymbol{S_1}^T & \boldsymbol{D_2} - \boldsymbol{S_2} \end{bmatrix} = \boldsymbol{D} - \boldsymbol{S},$$

and $\boldsymbol{L_0} + \lambda \boldsymbol{D_0}(\boldsymbol{\Omega_0^f} + \boldsymbol{\Omega_0^b}) \equiv \tilde{\boldsymbol{L}}$ is a positive definite matrix. In this study, the multiscale method of [4] is also applied.

The weighted mild constraint (6) has two advantages in the computational process of solving Eq.(10). First, it enforces the positive definiteness of the matrix $\tilde{\boldsymbol{L}}$, thereby making the computation more stable. Second, weight d_i normalizes the effect of the balancing parameter λ as follows. By using $s_{ii}^C = 0$, the i-th equation divided by d_i in Eq.(10) is

$$\left\{1 + \lambda(\Omega_i^f + \Omega_i^b)\right\} \alpha_i - \sum_{j \neq i} s_{ij}\alpha_j / d_i = \lambda \Omega_i^f - \{\boldsymbol{L_1}\boldsymbol{\delta}\}_i / d_i, \tag{11}$$

whereas the unweighted version would be written as

$$\left\{1 + \frac{\lambda}{d_i}(\Omega_i^f + \Omega_i^b)\right\} \alpha_i - \sum_{j \neq i} s_{ij}\alpha_j / d_i = \frac{\lambda}{d_i}\Omega_i^f - \{\boldsymbol{L_1}\boldsymbol{\delta}\}_i / d_i. \tag{12}$$

We focus on the balance between 1 and $\Omega_i^f + \Omega_i^b$ (affinity) which is approximately limited up to 1. In Eq.(12), λ would be divided by d_i, and the balance would be controlled not only by λ but also by d_i which is irrelevant to this balance. By adopting the weight d_i, λ directly acts on the balance in Eq.(11), and its effects are the same at any pixel in any images and are independent of d_i.

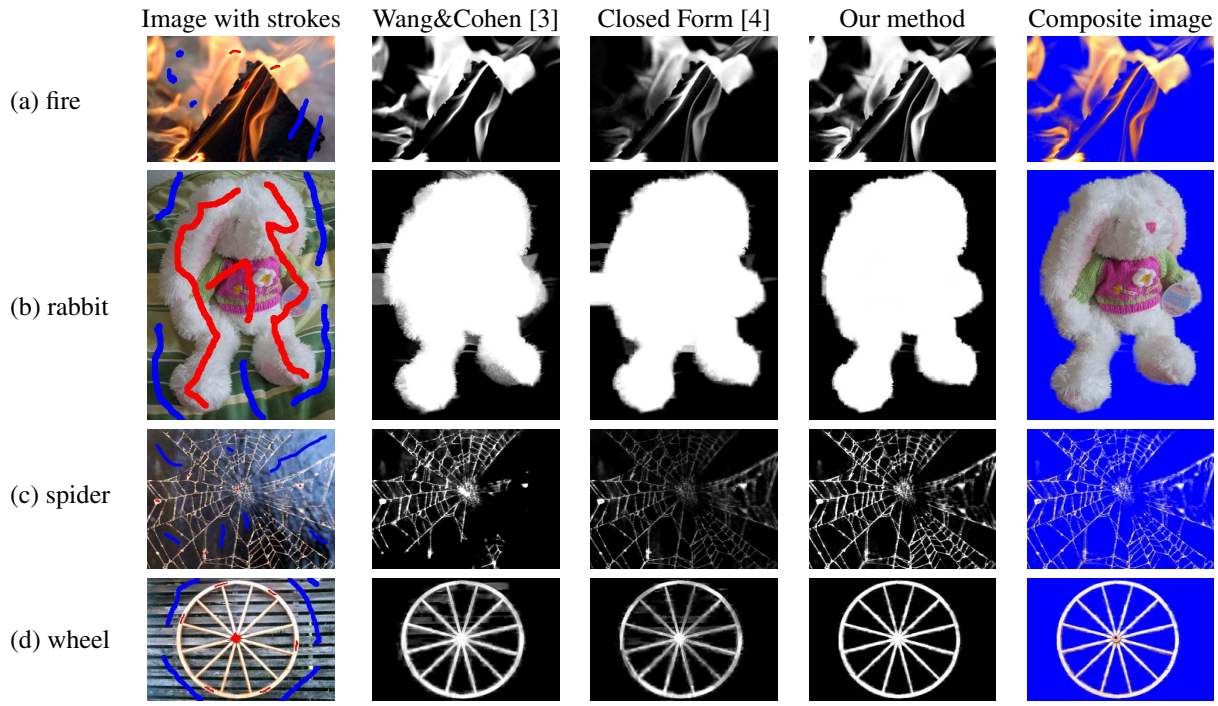

Fig. 2. Results of alpha mattes and composite images with extracted objects and a blue background.

4. EXPERIMENTAL RESULTS AND CONCLUSIONS

We applied the proposed method to various images and compared it with other state-of-the-art algorithms [3][4] by using the programs provided at their websites. We tuned the parameters of all the methods so that the appearance of the alpha matte seemed optimal. Fig.2 shows the resulting alpha mattes and composite images of four examples. Here, the upper two images also appeared in [3][4] with similar strokes. In spite of a few strokes, our method provides favorable results on the whole. Except for the *fire* image, the other algorithms result in erroneous alpha mattes because the backgrounds have similar colors to that of the foregrounds in Fig.2(b,c) and are highly textured in Fig.2(b,d). Note that our method favorably deals with two types of foreground objects –transparent objects (Fig.2(a,c)) and solid objects (Fig.2(b,d))– for which matting and segmentation approaches may be effective, respectively. For the 300×200 pixels image of Fig.2(a), 0.12 sec are required to calculate affinity and 2.65 sec to solve Eq.(10) by using a 2.6GHz CPU with 3GB RAM.

Taking into account of a review of previous studies, we have proposed a unified formulation for image matting in the framework of Q-IV, and the experiments have demonstrated its effectiveness. Two types of constraints are incorporated in our formulation: *hard* and *mild* constraints. The *hard* constraint corresponds to user inputs (strokes), as in the previous studies. The *mild* constraint is derived from the affinities based on the SVM classification result at each pixel. Our method produces favorable results for several kinds of images by using only a few user interactions (strokes).

5. REFERENCES

[1] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *CVPR*, 2001.

[2] L. Grady, T. Schiwietz, and S. Aharon, "Random walks for interactive alpha-matting," in *VIIP*, 2005.

[3] J. Wang and M. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *ICCV*, 2005.

[4] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *CVPR*, 2006.

[5] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," in *SIGGRAPH*, 2004.

[6] Y. Boykov and M.P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation for objects in n-d images," in *ICCV*, 2001.

[7] J. Shi and J. Malik, "Normalized cuts and image segmentation," *PAMI*, 22(8):888–905, 2000.

[8] C. Hayashi, "On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view," *Annal of the Institute of Statistical Mathematics*, 3(2):69–98, 1952.

[9] J. B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage Publications, Beverly Hills, CA, 1977.

[10] T. Hosaka, T. Kobayashi, and N. Otsu, "Image matting using SVM and neighboring information," in *VISAPP*, 2007.