Proceedings of the Seventh IASTED International Conference
VISUALIZATION, IMAGING, AND IMAGE PROCESSING
August 29–31, 2007, Palma de Mallorca, Spain
ISBN Hardcopy: 978-0-88986-691-1 / CD: 978-0-88986-692-8

AN AUTOMATIC EVALUATION METHOD OF IMAGE MATTES

Takumi Kobayashi and Hosaka Tadaaki and Nobuyuki Otsu
National Institute of Advanced Industrial Science and Technology, Japan
{takumi.kobayashi, hosaka-t, otsu.n}@aist.go.jp

ABSTRACT
In image matting, several algorithms that produce high-quality results have recently been proposed; however, certain parameters have to be manually determined in order to obtain favorable results, which requires significant user effort. In this paper, we propose a method of automatically evaluating alpha mattes based on either of the two criteria: mutual information or correlation coefficient between the matte and the image gradient fields, and selecting optimal parameter values. This drastically reduces the user effort required for tuning parameters in image matting. The experimental results using several matting algorithms for various images show that the automatically selected alpha mattes are similar (or the same) to the manually selected optimal ones.

KEY WORDS
Image Matting, Automatic Evaluation, Mutual Information, Correlation Coefficient

1 Introduction

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

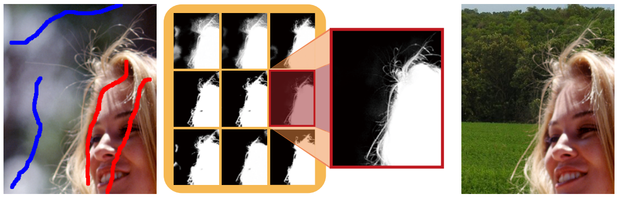
$$\boldsymbol{C}_i = \alpha_i \boldsymbol{F}_i + (1 - \alpha_i)\boldsymbol{B}_i, \tag{1}$$

where $\alpha_i \in [0, 1]$ represents the opacity; $\boldsymbol{C}_i$, the color vector in an image; and $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$, 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 and B in $\boldsymbol{C}_i$) and seven unknowns to estimate in Eq.(1). Several algorithms have been recently proposed to deal with the ill-posedness in the computer vision community. Unlike blue-screen matting [9], these algorithms have shown high-quality results without controlling the environment.

In these algorithms, some user interactions are required for indicating the foreground object to extract; these interactions also function as clues or constraints for solving the problem. One method for such user interactions is a *trimap* [8, 2, 4, 10], in which an image is segmented by a user into three regions: definitely foreground, definitely background, and unknown. The unknown region, in which the alpha values are estimated, is expected to possibly be a narrow strip for the sake of obtaining high-quality results. Another method for the user interactions, *strokes*, has recently been proposed for image matting [11]. As shown in Fig.1(a), a user draws two types of strokes: definitely foreground and background. The degree of the user interaction is significantly less than that of trimap. However, the region of alpha estimation is larger when using strokes, which makes the matting problem more difficult. Wang and Cohen [11] estimated the alpha values by iteratively propagating them from the strokes by using belief propagation. Levin *et al.* [7] transformed the above ill-posed problem into a closed-form expression under local smoothness assumptions on the foreground and background colors. Hosaka *et al.* [5] incorporated the discriminative information between foreground and background into the matting formulation of the MRF framework. Kobayashi *et al.* [6] have also proposed the unified formulation of alpha estimation which naturally incorporates the discriminative information in the closed form.

In a case that the estimated alpha matte is not satisfactory for a user, two approaches are usually employed for its improvement: repeating the alpha estimation by adding more clues and/or changing the parameters of the alpha estimation method. Without the appropriate parameters, even if a user adds some more clues (strokes) to modify the alpha matte, a favorable result might not be obtained. Therefore, the parameters first need to be appropriately determined, and then the user effort, *e.g.*, adding strokes, is required for further improving the alpha mattes, if necessary. In the alpha estimation, there are usually several parameters, such as regularization parameters, that need to be manually determined by a user. Although these parameters increase the degree of freedom in the formulation to deal with various images, it is an exhaustive task for a user to decide the optimal parameter values that produce the most favorable result by trial and error.

In this paper, we propose a method of automatically evaluating the alpha mattes. Our method searches the optimal alpha matte by evaluating all the mattes produced using various parameter values of the employed matting algorithm, as illustrated in Fig.1, which significantly reduces user effort required for tuning parameters. The evaluation criterion is based on the consistency of the relationship between the matte and the image gradient fields. Since the alpha mattes are evaluated after performing a matting algorithm, it can be integrated with any matting algorithms



(a) Input (b) Automatic selection (c) Composite

Figure 1. Once a user draws cue strokes for the foreground and the background in an input image (a), our method automatically selects the optimal alpha matte (b) among all the mattes produced using various parameter values. Figure (c) is a composite image with the extracted object and a new background.

as post-processing. The experimental results show that the evaluation method favorably selects the alpha mattes.

2 Automatic Evaluation of Alpha Mattes

Practically, there are several parameters in most matting algorithms. Since the parameters in the alpha estimation can not be appropriately determined in advance, they are usually selected on the basis of the observation of the resulting alpha mattes so as to produce the most favorable alpha matte for a user. In order to make a user free from tedious tuning parameters, the evaluation of the alpha mattes needs to be automated. Thus, we focus on how the user evaluates the alpha mattes produced using various parameter values, and then derive a criterion for the automatic evaluation of the alpha mattes. The criterion will enable the quantitative evaluation of the alpha mattes, which has not been performed in previous studies except for using artificial ground truth [7].

2.1 Considering the Evaluation by Users

Users can evaluate the alpha mattes according to their impressions, although it is not actually clear what these impressions are based on. However, the alpha mattes are probably evaluated on the basis of the top-down knowledge about the foreground object, and bottom-up information derived from comparing the alpha matte with the original image. The top-down knowledge requires the high-level cognition which is still one of the most difficult problems in computer vision. In this paper, we employ the bottom-up approach under the assumption that the comparisons between the *contours* of the foreground objects in an alpha matte and those in the original image play an important role in the evaluation. For example, when evaluating two alpha mattes, shown in Fig.2 (b,c), estimated from the original image (a) using different parameter values, alpha matte (c) would be selected as the favorable one by the user. In (b), some of the contours correspond to those in (a) (lower right window in (a,b)), whereas the other corresponding regions do not have the same contours (other windows in (a,b)). Such an inconsistency occurs in unsatisfactory alpha mattes. On the other hand, as to satisfactory result (c), most contours correspond to those in (a). In this paper, the criterion for evaluating the alpha mattes is derived from these observations. The key concept is that the contours in a satisfactory alpha matte must be shared with the original image. In image matting, however, it is not suitable to extract the contours and directly compare them. This is because the foreground objects are represented by their opacities in Eq.(1), and thereby the contours are not necessarily explicitly extracted as in the case of hard segmentation, especially for the transparent objects. Therefore, we focus on the fact that steep gradients exist around the contours in both the alpha matte and the original image. On this basis, the above concept implies that the pixels at which steep gradients are detected in the alpha matte should consistently have steep gradients in the corresponding original image. The criterion is defined below based on the relationship between the matte and the image gradients.

2.2 Evaluation Criterion

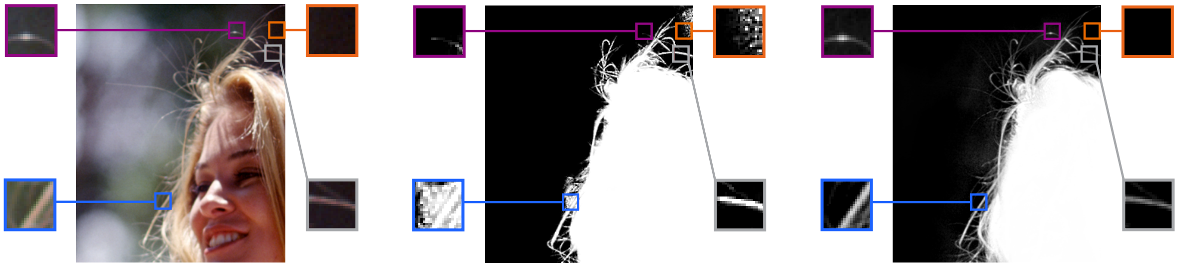
On the basis of the above observations, a criterion is constructed so as to evaluate the consistency of the relationship between gradients in the alpha matte and in the original image. In this paper, we propose the following two types of criteria: *mutual information (MI)* and *correlation coefficient (CC)*. In these criteria, it is expected that the matte gradients should be positively correlated with the image gradients and the inconsistencies such as those in Fig.2(b) are not permitted. Here, the following notations are used for the definition of the criteria. The gradient at pixel i is represented by the magnitude S_i^α and the direction θ_i^α in an alpha matte and by S_i^o and θ_i^o in the original image, respectively. The direction ranges from 0 to π and thus it is cyclic in the range $\theta \in [0, \pi)$, which is slightly difficult to be dealt with when calculating the criteria. Since only the relationship between θ_i^α and θ_i^o is important in the evaluation, we transform the value of θ_i^o such that $|\theta_i^o - \theta_i^\alpha| \leq \pi/2$. $\mathcal{P}_E^\alpha$ denotes the set of pixels having gradients that are steeper than the threshold ξ in the alpha matte: $\mathcal{P}_E^\alpha = \{i | S_i^\alpha > \xi\}$. These pixels are mostly located around contours in the alpha matte to be evaluated and are focused on in the following criteria.

Mutual Information Criterion (MI): The MI between the distributions of $(S^\alpha, \theta^\alpha)$ and (S^o, θ^o) in $\mathcal{P}_E^\alpha$ is calculated as

$$\mathrm{MI} = I(S^\alpha; S^o) + I(\theta^\alpha; \theta^o) \tag{2}$$

$$I(S^\alpha; S^o) = \iint p(S^\alpha, S^o) \log \frac{p(S^\alpha)p(S^o)}{p(S^\alpha, S^o)} dS^\alpha dS^o$$

$$I(\theta^\alpha; \theta^o) = \iint p(\theta^\alpha, \theta^o) \log \frac{p(\theta^\alpha)p(\theta^o)}{p(\theta^\alpha, \theta^o)} d\theta^\alpha d\theta^o,$$



(a) (b) (c)

Figure 2. Evaluation of alpha mattes. (a) Original image: the square windows are associated with those in alpha mattes (b,c). (b) Unsatisfactory alpha matte: only the right-down window is satisfactory while the other windows are not. (c) Satisfactory alpha matte: most contours are consistent with those in the original image.

where $p(\cdot)$ indicates a probability defined by samples (pixels) of $\mathcal{P}_E^\alpha$. As to the probability of magnitude S, each sample $i \in \mathcal{P}_E^\alpha$ has two variates (S_i^α, S_i^o) and all of samples form the probability distribution of 2-dimension, $p(S^\alpha, S^o)$, and 1-dimension, $p(S^\alpha)$ and $p(S^o)$. The same holds for that of direction θ. In Eq.(2), we assume that the magnitude S is independent of the direction θ. The value of MI increases as matte gradients depend on image gradients. These dependencies between the gradients in the two images are evaluated in this criterion and inconsistent (independent) relationships cause a decrease in MI. Even nonlinear relationships can be captured in this criterion.

Correlation Coefficient Criterion (CC): The CC between samples $\{S_i^\alpha, \theta_i^\alpha\}$ and $\{S_i^o, \theta_i^o\}$ in $\mathcal{P}_E^\alpha$ is defined as

$$\mathrm{CC} = R_S + R_\theta \tag{3}$$

$$R_S = \frac{\sum_{i\in\mathcal{P}_E^\alpha}(S_i^\alpha - \bar{S^\alpha})(S_i^o - \bar{S^o})}{\sqrt{\sum_{i\in\mathcal{P}_E^\alpha}(S_i^\alpha - \bar{S^\alpha})^2}\sqrt{\sum_{i\in\mathcal{P}_E^\alpha}(S_i^o - \bar{S^o})^2}}$$

$$R_\theta = \frac{\sum_{i\in\mathcal{P}_E^\alpha}(\theta_i^\alpha - \bar{\theta^\alpha})(\theta_i^o - \bar{\theta^o})}{\sqrt{\sum_{i\in\mathcal{P}_E^\alpha}(\theta_i^\alpha - \bar{\theta^\alpha})^2}\sqrt{\sum_{i\in\mathcal{P}_E^\alpha}(\theta_i^o - \bar{\theta^o})^2}},$$

where $\bar{\cdot}$ denotes the mean in $\mathcal{P}_E^\alpha$. We simply sum up the two correlation coefficients, R_S and R_θ, of magnitude S and direction θ in Eq.(3) in a manner similar to that in Eq.(2). CC evaluates the correlations between the matte and the image gradients. It captures linear relationships, and the inconsistent relationships (negative correlations) decrease the evaluation score.

Although MI adapts to various relationships due to its nonlinearity, it could possibly allow even negative correlations. Therefore, it is unclear which of the two criteria performs better in evaluating the alpha mattes. The experimental results in Sec.4, however, slightly clarify what kind of image each criterion tends to be suitable for.

2.3 Selection of Optimal Parameters

In our method, the optimal alpha matte is searched by the automatic evaluation of the alpha mattes as follows. First, the alpha mattes are produced by any matting algorithms using various parameter values, and then they are evaluated respectively in the subsequent procedure. Second, by calculating gradients in the alpha matte and thresholding the magnitudes, the pixel locations of the steep gradients, $\mathcal{P}_E^\alpha$, are detected. We also calculate the magnitude S_i^o and the direction θ_i^o of the gradient at pixel $i \in \mathcal{P}_E^\alpha$ in the original image. The gradients are calculated using color channels as in [3]. Third, the alpha matte is evaluated based on either the MI or the CC criterion using the relationships between $(S^\alpha, \theta^\alpha)$ and (S^o, θ^o). Finally, the alpha matte with the highest evaluation score is selected.

3 Relationship to Previous Works

In [7], the following cost function is minimized with respect to $\boldsymbol{\alpha}$,

$$J(\boldsymbol{\alpha}) = \min_{a,b} \sum_{j\in\text{Image}} \sum_{i\in w_j} (\alpha_i - a_j C_i - b_j)^2 + \epsilon a_j^2, \tag{4}$$

where w_j indicates the local window centered at pixel j and ϵ is a regularization parameter. Although this is derived from Eq.(1), it also means maximization of the normalized correlation between alpha values α and pixel values C within each *local* windows w_j in case of $a > 0$. In the proposed criterion, we maximize the normalized correlation (correlation coefficient) between matte gradients and image gradients on the contours in the alpha matte. The criterion evaluate the *global* consistency of the object extracted by matting algorithm.

Apostloff *et al.* [1] also adopted the concept of the relationship between the matte and the image gradients. They learnt the model of the relationship ($p(\nabla\alpha|\nabla \boldsymbol{C})$) from the reliable alpha mattes produced by blue-screen matting. The model, however, could be varied in every image and its construction is also an exhaustive task. In this paper, we measure the consistency of the relationships in each alpha matte by using bottom-up approach without employing such models. The model construction in [1] also suggests the validness of the consistency.

Sun *et al.* [10] focused on the matte gradient field and solved the associated Poisson equations, which are based on the following equations:

$$\nabla \alpha_i = \frac{1}{\boldsymbol{F}_i - \boldsymbol{B}_i}\{\nabla \boldsymbol{C}_i - \boldsymbol{D}_i\}, \quad (5)$$

where $\boldsymbol{D}_i = (\alpha_i \nabla \boldsymbol{F}_i + (1-\alpha_i)\nabla \boldsymbol{B}_i)$. In global Poisson matting, the matte gradient $\nabla \alpha_i$ is assumed to be linearly related to the image gradient $\nabla \boldsymbol{C}_i$, where $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$ are smooth, i.e., $\boldsymbol{D}_i \approx \boldsymbol{0}$. In local Poisson matting, $\boldsymbol{D}_i$ was modeled in several ways for the region that $\boldsymbol{F}_i$ and $\boldsymbol{B}_i$ are not smooth. In such a case, the relationships between the matte and the image gradients are nonlinear as in the case of Eq.(5) with $\boldsymbol{D}_i \neq \boldsymbol{0}$. As above, the two types of relationships naturally appear in the image matting. Our criteria can respectively deal with these relationships: MI for nonlinear and CC for linear.

4 Experimental Results

We evaluate the alpha mattes produced by state-of-the-art matting algorithms [11, 7, 5, 6]. These algorithms have several parameters to be determined by a user and employ the user interactions of strokes type. It should be noted that for the algorithms of [11, 7] we utilize the programs provided at their websites and can change parameters in their programs. The parameters in these algorithms are briefly described as follows, and refer to each paper for more details.

Wang and Cohen [11]: There is one major parameter λ_s (sensitivity) which balances the two terms in the following cost function,

$$\sum_i V_d(\alpha_i) + \lambda_s \sum_{(i,j)\in\text{Neighbor}} V_s(\alpha_i, \alpha_j), \quad (6)$$

where V_d and V_s are data energy and smoothness energy, respectively. The other (minor) parameters are set to the appropriate values recommended by the authors. The range of parameter is that $\lambda_s \in [0, 1]$ with 20 steps, and then 20 alpha mattes are produced.

Levin *et al.* [7]: There are three parameters: regularization parameter ϵ in Eq.(4), the number of multiscale layers M and the threshold for alpha values T_α in the coarse-to-fine scheme. The range of parameters is that $\epsilon \in \{1e^{-5}, 1e^{-7}\}, M = \{1,2,3,4,5,6\}, T_\alpha \in \{0.05, 0.12, 0.30\}$. 36 alpha mattes are produced.

Hosaka *et al.* [5]: There are two major parameters λ_M, λ_D which balances the three terms in the following cost function,

$$\lambda_M \sum_i U_M(\alpha_i) + \sum_{(i,j)\in\text{Neighbor}} U_S(\alpha_i, \alpha_j) + \lambda_D \sum_i U_D(\alpha_i) \quad (7)$$

where U_M, U_S, U_D are matting term, smoothing term and data term, respectively. The range of parameters is that $\lambda_M, \lambda_D \in [0, 2]$ with 23 steps, and then 529 alpha mattes are produced.

Kobayashi *et al.* [6]: This method is extended from [7] by adding second term in the following cost function,

$$\sum_{(i,j)\in\text{Neighbor}} s_{ij}(\alpha_i - \alpha_j)^2 + \lambda \sum_i d_i\{\Omega_i^b \alpha_i^2 + \Omega_i^f (1-\alpha_i)^2\} \quad (8)$$

where s_{ij} indicates similarity between pixel i and j, and $d_i = \sum_j s_{ij}$. The second term (Ω^b, Ω^f) is derived from the discrimination for fore/background at every pixel. Thus, in addition to three parameters (ϵ, M, T_α) in [7], there are two parameters: the number of neighboring pixels N used for the discrimination and balancing parameter λ in Eq.(8). The range of these parameters is that $N \in \{1, 27\}, \lambda \in [0, 0.5]$ with 33 steps, and $M \in \{1,2,3,4,5,6\}$ while ϵ and T_α are fixed as $1e^{-5}$ and 0.05, respectively. In total, 396 alpha mattes are produced.

Our method of automatic evaluation is tested in several images, and the selected alpha mattes for each matting algorithm are shown in Fig.4. In this experiment, the point is not the comparison among methods but the quality of the selected alpha matte in the parameter space. In the evaluation, both criteria, CC and MI, are simultaneously applied and two alpha mattes are chosen as a result. Since the characteristics of the criteria are different as discussed below, it is easy for a user to select the better one of the two by comparing them. The caption in Fig.4 indicates the criterion used for the better matte shown in the figure. The results except Fig.4(i,x) are favorable and much similar (or the same) to the manually selected optimal ones. As to the result of (i,x), the method [5] could not produce completely favorable matte in its parameter range. Even though there is inherently no optimal parameter, these mattes (i,x) are actually better ones in the parameter range of [5]. Therefore, it is found that our evaluation method favorably selects the alpha mattes which are also chosen in case of manual selection. For demonstrating the effectiveness of the criteria, the alpha mattes which are *not* selected due to its low evaluation score are also shown in Fig.3. In the parameter range, various alpha mattes are produced and not-selected ones are apparently worse than the selected one. Thus, it is shown that the proposed evaluation criteria are useful for automatic selection and agree with human impression.

The two criteria tend to specialize different types of images: CC works particularly well for transparent objects such as fire, face and peacock in Fig.4, while MI appears to be suitable for solid objects such as dog in Fig.4. Gradients change gradually in transparent objects while they change drastically around the borders in solid objects. These tendencies are associated with the linear and nonlinear correlations of gradients for which CC and MI are suitable, respectively. This distinction, however, is not clearly but moderately appeared, particularly for the image of child in Fig.4 which includes both solid region at the side of face

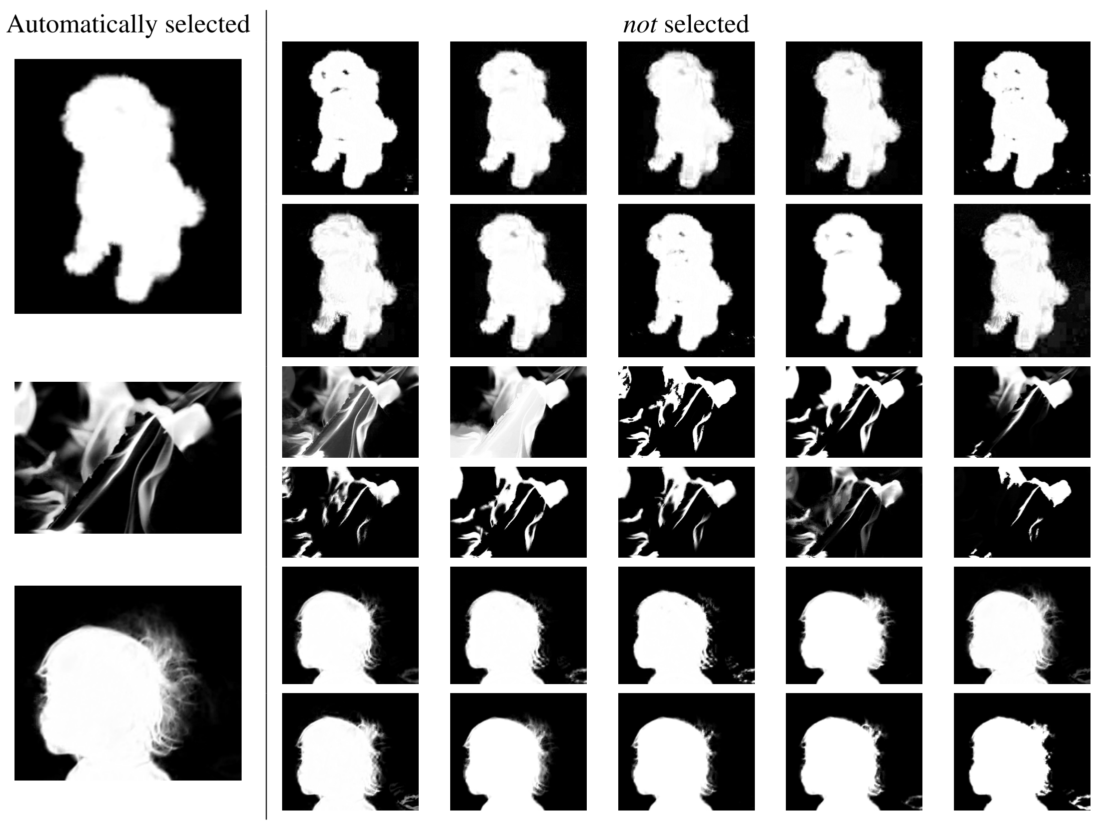


Figure 3. *Not* selected alpha mattes. These alpha mattes are apparently worse results.

and transparent region of hair. Computational time is dependent on that of employed matting algorithm, i.e., the parameter range and the processing time for one image, because the evaluation, of which task is only calculation of gradients, requires little computational cost.

5 Conclusion

In this paper, we have proposed a method for automatically evaluating alpha mattes and selecting optimal parameter values for image matting. Since our method is applied as post-processing after alpha estimation, it can be integrated with any alpha estimation algorithms including parameters. The best performance of the alpha estimation method is provided by automatically searching for the optimal results; this leads to a significant reduction in user effort for tuning parameters. Our key contribution is the construction of two types of criteria for the evaluation: mutual information and correlation coefficient. In the experiments using state-of-the-art matting algorithms for various images, the effectiveness of the proposed method is demonstrated.

References

[1] N. Apostoloff and A. Fitzgibbon. Bayesian video matting using learnt image priors. In *Proc. CVPR*, pages 407–414, 2004.

[2] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Proc. CVPR*, pages 264–271, 2001.

[3] A. Cumani. Edge detection in multispectral images. *CVGIP: Graphical Models and Image Processing*, 53(1):40–51, 1991.

[4] L. Grady, T. Schiwietz, and S. Aharon. Random walks for interactive alpha-matting. In *Proc. VIIP*, pages 423–429, 2005.

[5] T. Hosaka, T. Kobayashi, and N. Otsu. Image matting using svm and neighboring information. In *Proc. International Conference on Computer Vision Theory and Applications*, pages 344–349, 2007.

[6] T. Kobayashi, T. Hosaka, and N. Otsu. Image matting in the framework of quantification iv. In *Proc. ICIP*, 2007.

[7] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. In *Proc. CVPR*, pages 61–68, 2006.

[8] M. Ruzon and C. Tomasi. Alpha estimation in natural images. In *Proc. CVPR*, pages 18–25, 2000.

[9] A. Smith and J. Blinn. Blue screen matting. In *Proc. ACM SIGGRAPH*, pages 259–268, 1996.

[10] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum. Poisson matting. In *Proc. ACM SIGGRAPH*, pages 315–321, 2004.

[11] J. Wang and M. Cohen. An iterative optimization approach for unified image segmentation and matting. In *Proc. ICCV*, pages 936–943, 2005.

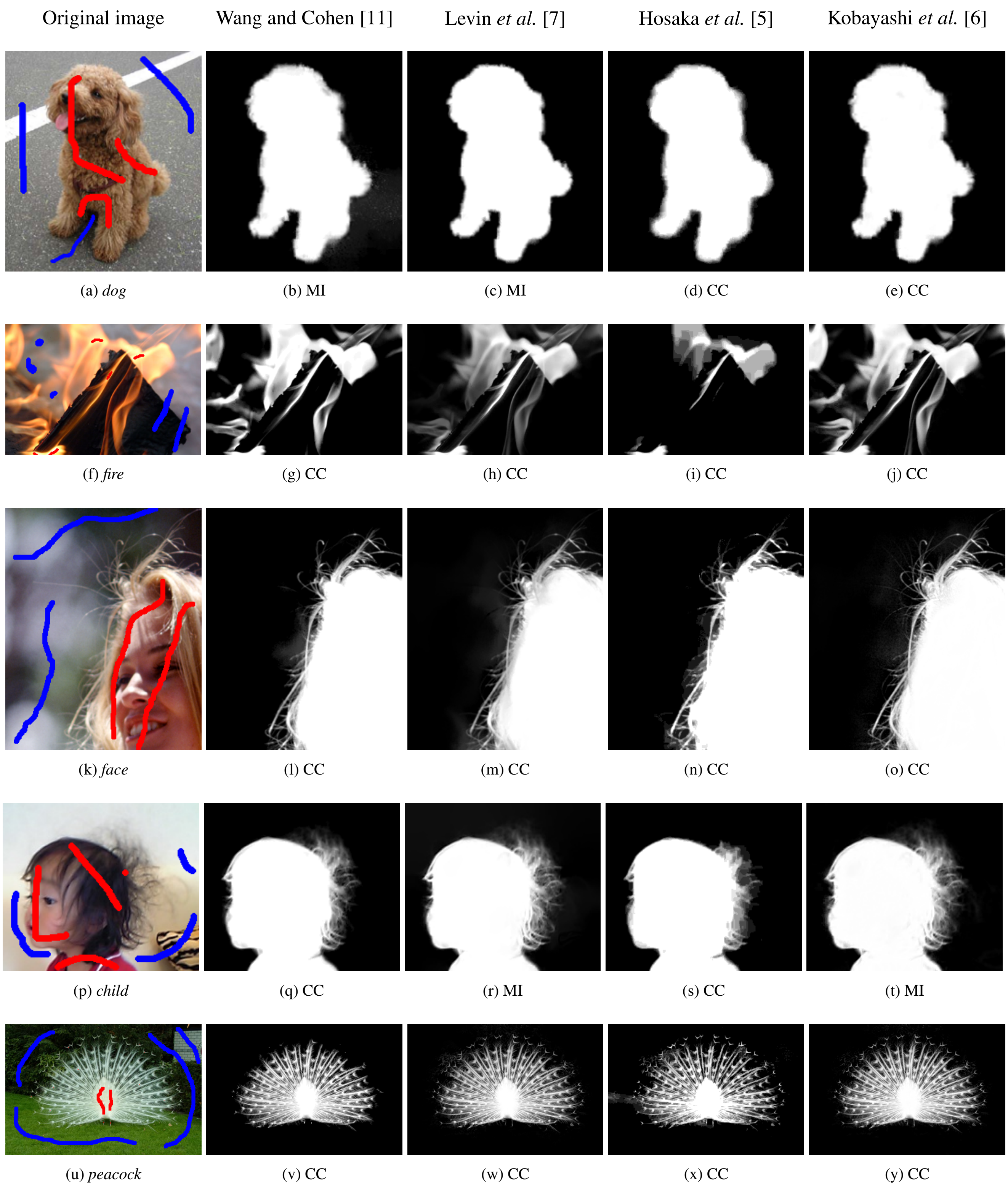


Figure 4. Automatically selected alpha mattes. The caption of each alpha matte means which criterion (CC or MI) is used to automatically select the alpha matte.