Automatic Trimap Generation and Consistent Matting for Light-Field Images

Donghyeon Cho, *Student Member, IEEE,* Sunyeong Kim, Yu-Wing Tai, *Senior Member, IEEE,* and In So Kweon, *Senior Member, IEEE*

Abstract—In this paper, we introduce an automatic approach to generate trimaps and consistent alpha mattes of foreground objects in a light-field image. Our method first performs binary segmentation to roughly segment a light-field image into foreground and background based on depth and color. Next, we estimate accurate trimaps through analyzing color distribution along the boundary of the segmentation using guided image filter and KL-divergence. In order to estimate consistent alpha mattes across sub-images, we utilize the epipolar plane image (EPI) where colors and alphas along the same epipolar line must be consistent. Since EPI of foreground and background are mixed in the matting area, we propagate the EPI from definite foreground/background regions to unknown regions by assuming depth variations within unknown regions are spatially smooth. Using the EPI constraint, we derive two solutions to estimate alpha when color samples along epipolar line are known, and unknown. To further enhance consistency, we refine the estimated alpha mattes by using the multi-image matting Laplacian with an additional EPI smoothness constraint. In experimental evaluations, we have created a dataset where the ground truth alpha mattes of light-field images were obtained by using the blue screen technique. A variety of experiments show that our proposed algorithm produces both visually and quantitatively high-quality alpha mattes for light-field images.

Index Terms—Image Matting, Light-Field Image, Trimap.

1 Introduction

Image Matting aims to extract soft and accurate alpha matte of foreground given a trimap of an image. Generally, colors of an image can be expressed as a linear combination of foreground and background colors as follows:

$$I = \alpha F + (1 - \alpha)B, \tag{1}$$

where F, B and α represent the foreground, the background, and the mixing coefficients, respectively. Since most matting algorithms were developed for single image matting, it is less effective when facing multiple input images, e.g. multiple sub-images of a light-field image, where consistent alpha mattes across the multiple images are necessary. In this paper, we introduce a new image matting algorithm targeting for a light-field image.

A light-field image consists of $m \times n$ sub-images where each sub-image were captured from slightly different perspectives. The correlation among the sub-images are encoded in the epipolar plane image (EPI), and the estimated alpha mattes across sub-images also need to follow the EPI constraint. Otherwise flickering artifacts will appear when moving an interpolated view point from one sub-image to another sub-image.

- *D. Cho and I.S. Kweon are with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. E-mail: cdh12242@gmail.com, iskweon@kaist.ac.kr*
- *S. Kim is with the Perples, Inc, Seoul, Korea. E-mail: harharr@gmail.com*
- *Y-W. Tai is with SenseTime, HongKong. E-mail: yuwing@gmail.com*
- *Corresponding author: I.S. Kweon.*

Manuscript received XXXX XX, XXXX; revised XXXX XX, XXXX.

By Eq. (1), image matting is an ill-posed problem because the number of unknowns is more than the number of equations that can be derived from a single image. State-of-the-art matting algorithms can be categorized into two groups: color sampling based, and alpha propagation based methods. The color sampling based methods [1], [2], [3], [4], [5], [6], [7] sample foreground and background colors from the known regions, *i.e.* the definite foreground and the definite background regions, to estimate alpha mattes within the unknown region. The alpha propagation based approaches [8], [9], [10], [11], [12], [13], [14] assumes local/nonlocal smoothness of alpha values and propagate alpha values from the known regions to the unknown regions.

In the light-field image matting problem, although the number of input images have increased, the number of unknown have also increased which makes it also an ill-posed problem. However, because of the EPI correlation across the sub-images, we can sample foreground and background colors across sub-images. Even if a true color sample in a sub-image is missing, we can still reliably estimate the true color samples from another sub-image. This allows us to achieve better performance than existing color sampling matting techniques. Using the EPI constraint, we can propagate alpha values not only from the known regions to the unknown regions within a sub-image, but also along EPI of alpha mattes across sub-images. This provides an accurate and consistent alpha estimation across sub-images. In addition, if we consider the foreground is the closest object to a camera, we can utilize the EPI to estimate depth map, and identify foreground objects automatically. As demonstrated in our paper, our approach can automatically generate a reliable trimap based on colors and depth information.

Our approach consists of two major components: auto-

matic trimap generation, and consistent alpha matte estimation. In the trimap generation, we first estimate a binary segmentation of foreground and background using colors and depth information. The binary segmentation does not need to be very accurate, since our approach estimates unknown regions with adaptive analysis of color distributions along the segmentation boundary. In the consistent alpha matte estimation, we utilize the EPI constraint to propagate known color samples from one sub-image to another sub-image. A closed form solution is derived for alpha estimation, when two or more color samples along the same epipolar line of alpha are known. We have also derived a solution, when color samples along epipolar line are unknown. As demonstrated in our experimental results, our approach reduces weaknesses and maximize strengths of both kinds of image matting techniques.

We evaluate and compare performance of our proposed algorithm and state-of-the-art image matting algorithms. To quantitatively compare the performance, we created a new light-field matting dataset where the ground truth alpha mattes are obtained by using the blue screen matting procedures introduced in [15]. Our evaluations show that our algorithm produces both visually and quantitatively high-quality alpha mattes for light-field images, and have outperformed existing matting algorithms in term of both accuracy and consistency.

A shorter version of this work appeared in [16]. This paper extends our previous work with further technical details of our implementation, and additional experimentations. The algorithm for automatic trimap generation is newly introduced. To summarize, our contributions are as follows:

1) We introduce a new algorithm to generate trimaps of a light-field image automatically.
2) We introduce a method to construct color sample sets across multiple sub-images, and utilize the EPI constraint to estimate consistent alpha mattes along epipolar lines.
3) We introduce EPI smoothness constraint in the matting Laplacian to enhance consistency of estimated alpha mattes across sub-images.
4) A dataset with ground truth alpha mattes is created to quantitatively evaluate performance of matting algorithms on light-field images.

2 Related Work

We review previous works that are the most relevant to our work. In particular, we discuss the works related to the two categories of image matting and the works related to light-field image processing.

As aforementioned, most image matting techniques can be categorized into color sampling based and alpha propagation based methods. The color sampling based methods [1], [2], [3], [4], [5], [6], [7] solve the matting problem by using color samples from the known foreground and background regions to estimate alpha mattes in unknown regions. In [1], the Bayesian matting by Chuang *et al.* analyzes colors of unknown pixels using local color distribution by statistical methods. Robust Matting [2] collects color samples with respect to the color composite equation and are spatially close to the unknown pixels. Shared matting [4] and weighted color and texture matting [6] find the best samples by combining spatial, photometric, and probabilistic information measured by color and texture, respectively. In [5], He *et al.* proposed the global sampling matting which uses all color samples in the known regions to find the best combination of foreground and background samples for matte estimation. Recently, Shahrian *et al.* [7] proposed the comprehensive sampling matting which uses Gaussian Mixture Model (GMM) to cover all color variations in the foreground and background regions of an image for accurate alpha matte estimation. Once the best pairs (F, B) of color samples are selected, α for each pixel is computed as:

$$\hat{\alpha} = \frac{(I - B) \cdot (F - B)}{||F - B||^2}. \quad (2)$$

The alpha propagation based approaches [8], [9], [10], [11], [12], [13], [14] analyze statistical correlation among pixels to propagate alpha values from the known regions to the unknown regions. The Poisson Matting by Sun *et al.* [8] estimates an alpha matte by solving the Possion equation to reconstruct an alpha matte from gradients subject to the boundary condition of alpha matte in the known regions. Levin *et al.* [9] introduced the color line model and propose the matting Laplacian to solve the matte estimation problem in a closed form. This work is later extended by He *et al.* [11] whom proposed the large kernel matting for high resolution image matting. Based on the nonlocal principle, Lee and Wu [12] introduced the nonlocal matting which propagate alpha values across nonlocal neighbor of a pixel. This work is later extended by Chen *et al.* [13] whom proposed the KNN matting which propagates alpha across the k nearest nonlocal neighbors. Work by Chen *et al.* [14] combines local and nonlocal smoothness prior for alpha propagation which achieved the state-of-the-art performance in natural image matting. More recently, Cho *et al.* [17] introduces a deep learning approach to combine results from closed form matting, and KNN matting which achieves the most state-of-the-art results in natural image matting.

Several works [18], [19], [20], [21], [22], [23] directly deal with a trimap to estimate unknown regions automatically. McGuire *et al.* [19] utilize multiple synchronized video frames, each with different amount of defocus, to estimate video trimaps automatically. Joshi *et al.* [20] presented a high quality video matting method using a camera array. They construct a trimap automatically based on variance measurements across video frames from different perspectives. Sun *et al.* [24] extract alpha mattes using a pair of flash/no-flash images. They estimate a trimap based on two-pass thresholding method using a flash-only image. Wang *et al.* [18] use depth information from a ToF camera to estimate a trimap automatically. Rhemann *et al.* [21] suggested an interactive trimap segmentation via energy minimization. He *et al.* [22] iteratively refine a binary mask and a depth map of RGB-D images in alternating manner. Kim *et al.* [23], [25] considers the geometry constraint across multi-view images to extract foreground object automatically with fractional boundaries.

In light-field image processing, since Ng *et al.* [26] introduced the first prototype of micro lens array light-field camera, a lot of follow up works have been proposed. Cho

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

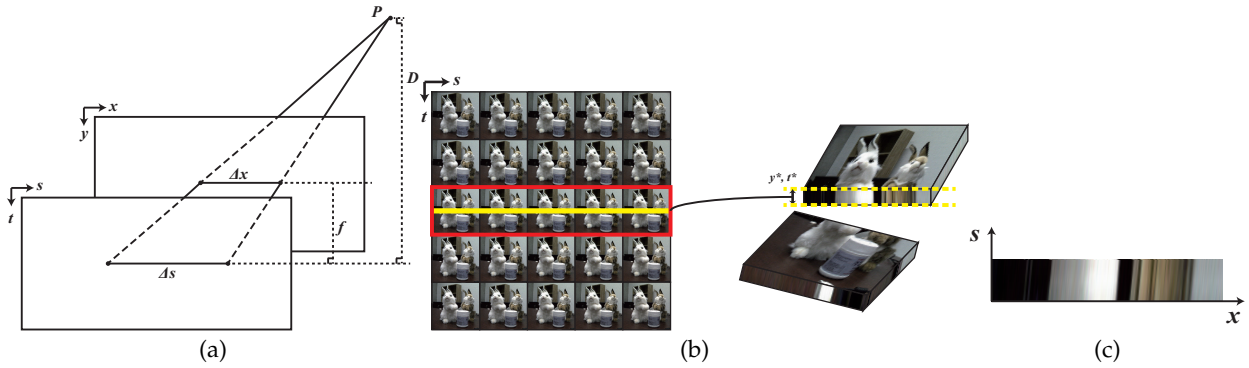


Fig. 1. (a) Parallel plane representation of a light-field image. (b) The multiple sub-images of a light-field image after decoding. (c) After stacking the images within the yellow line region in (b), we have an epipolar plane image in $x-s$ plane with fixed y,t.

et al. [27] presented decoding processes which produce 4D light-field images from a raw image and a learning based super-resolution technique directly on hexagonal arrangement of a raw image. The work by Dansereau *et al.* [28] estimates a depth map of the corresponding elements in a scene using gradient vector. Bishop and Favaro [29] estimate depth map by evaluating aliasing across multiple views. Wanner *et al.* [30], [31] use epipolar plane images to estimate depth map with consideration of global and local consistent. The work by Goldluecke and Wanner [32] computes depth maps by using the local derivative constraint with a convex prior derived from a 4d light-field image. The work by Jeon *et al.* [33] constructs cost volume using the phase shift theorem followed by multi-label optimization to estimate depth map accurately. Also, Wanner and Goldluecke [34] suggest reconstruction method for reflective and transparent surfaces from 4D light-field images. Recently, Fiss [35] *et al.* model a 4D light-field image as a 4D background light-field with 2D spatially varying color layer, separate foreground layer from background using matting.

Compare our work with previous works, as far as we are aware, this is the first work which seriously addresses the matting problem in light-field images with automatic trimap generation. Unlike previous image matting methods, our proposed method utilizes depth and color information in two steps to estimate a trimap without user inputs. A binary segmentation is first performed to obtain rough boundary between foreground and background. Then, a trimap is generated by analyzing color distribution of boundary regions using the KL-divergence measures to cover various object structures. In terms of alpha matting algorithm, our method takes color samples along the EPI correspondences which significantly reduces ambiguities in color sample selection. As a post-processing, estimated alpha mattes are refined by the multi-image Laplacian with an additional smoothness constraint. For evaluation, we have created a dataset of two types of light-field cameras: a Lytro, and a lab-made light-field camera.

3 EPI in Light-Field Images

3.1 The EPI constraint

A light-field image is typically represented as a 4D function, $L(x, y, s, t)$, which records the intensity of a light ray passing through two parallel planes, $x-y$ and $s-t$ planes, in the 3D space as illustrated in Fig. 1(a). To capture a light-field image, one can use a camera array or a consumer level micro lens light-field camera, e.g. Lytro [36]. After decoding, we can obtain multiple sub-images where each sub-image represents image captured from slightly different perspective as illustrated in Fig. 1(b).

Since a light-field image captures light rays in the 3D space, a light ray from an object at different distance from camera would pass through the two parallel image plane at different angle. This relationship is captured in the EPI of a light-field image. For instance, if we fixed the index of y and t in $L(x, y, s, t)$, we can plot the EPI of $x-s$ plane as illustrated in Fig. 1(c). Mathematically, we can derive [30], [37], [38]:

$$\Delta x = \frac{f}{D}\Delta s, \tag{3}$$

where f is the distance between the two parallel image planes and D is the distance of a object from a camera as illustrated in Fig. 1(a). Using Eq. (3), we can obtain pixel correspondences across sub-images in x-direction by measuring the image gradients in the EPI of $x-s$ plane. Similarly, we can obtain pixel correspondences in y direction through measuring the image gradients in the EPI of $y-t$ plane. To generalize EPI correspondences at different directions (not only horizontal and vertical directions), we utilize the shearing introduced by Ng *et al.* [26]:

$$L_\gamma(x, y, u, v) = L_0(x + u(1 - \frac{1}{\gamma}), y + v(1 - \frac{1}{\gamma}), u, v), \tag{4}$$

where γ, L_0, and L_γ are a shearing variable, an input light-field image, and an sheared light-field image by γ, respectively. For each depth plane in the 3D world, there is a shearing value γ^* which makes its regions in-focus by digital refocusing:

$$I_{\gamma^*}(x, y) = \sum_{u,v} L_{\gamma^*}(x, y, u, v), \tag{5}$$

where I_{γ^*} is a refocused image. Thus, γ^* can be considered as EPI gradient of the 4D light-field space. When $I_{\gamma^*}(x^*, y^*)$ is in-focus, $L_{\gamma^*}(x^*, y^*, u, v)$ are correspondence pixels at spatial coordinate (x^*, y^*). Eq. (4) describes all directional corresponding rays of pixels at certain depth. Since colors of the corresponding pixels come from the same light ray in 3D, the estimated foreground/background colors as well as the alpha values are expected to be the same across the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

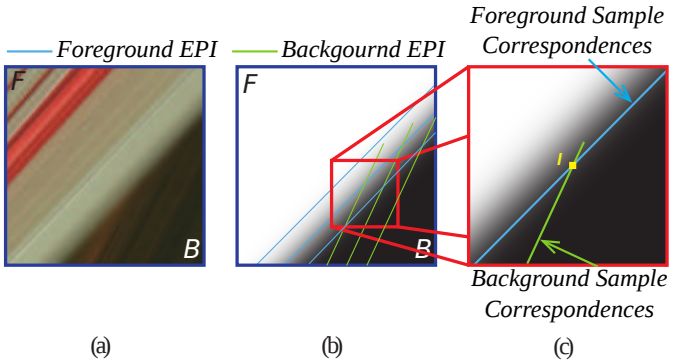


Fig. 2. (a) EPI of matting boundary. (b) Alpha matte of the EPI image. (c) Foreground and background color sample correspondences of a pixel in the matting area.

corresponding pixels. This defines the EPI constraint across the multiple sub-images of a light-field image.

3.2 Color Sample Correspondences in EPI

Using the EPI constraint, we can define pixel correspondences across the multiple sub-images and expect the color values as well as the alpha mattes along the EPI correspondences to be identical. In practice, because of the mixing effect in matting areas, the EPI constraint may not hold along the matting boundary since the measured intensity is the result of alpha blending of two light rays from different direction as illustrated in Fig. 2(a).

In order to utilize the EPI constraint, we need to estimate EPI gradient. Even though EPI gradient and depth estimation is not coverage of this paper, it is critical to performance of our works. If depth information is not reliable and noisy, finding corresponding pixels along with EPI line is not feasible. In our implementation, we first use depth estimation method in [39] and then apply a median filter to the estimated EPI. For matting boundary, we assume the EPI of foreground and background are spatially smooth in the unknown region of a trimap. The depth of foreground and background within the unknown region are similar to the depth of foreground and background in the known region. Thus, we can propagate the EPI constraint from the known region to the unknown region through extrapolation by solving the Poisson equation by setting zero gradient of EPI subjects to the boundary constraint of the estimated EPI along the trimap boundary. Fig. 2(c) illustrates the color sample correspondences of a pixel in a matting area using the propagated EPI constraint. Note that the foreground and the background color sample correspondences are defined differently since the foreground and background EPI have different light ray direction. We will discuss how to use the EPI color sample correspondences to select color samples for better alpha matte estimation in Sec. 5. Also in the case of multiple depth in the foreground, multi-label assignment technique for 4D light-field [31] will be helpful to find EPI correspondences.

3.3 Assumption

We assume foreground and background are located at different depth from camera such that the foreground and the background color sample correspondences are misaligned as illustrated in Fig. 2(c). If foreground and background are located at the same (or very closed) depth, the observed intensity across multiple sub-images will be identical. In such case, there are few benefits from a light-field image. Similar to conventional single image matting, user inputs are required to obtain definite foreground and background samples. Indeed, if foreground and background EPI are in the same direction, *i.e.* they are in the same depth, there is actually no mutual information across sub-images which can be used to assist image matting. Therefore, in this paper, we focus on a case that EPI gradients of foreground and background are quite different.

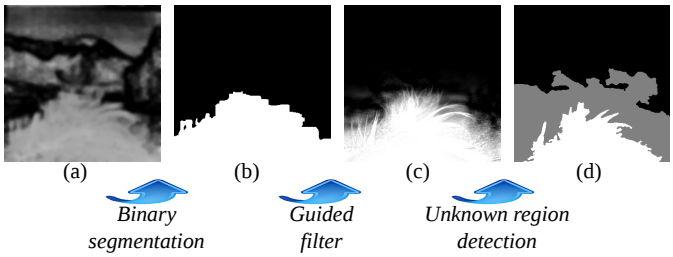


Fig. 3. (a) Estimated depth map from [39]. (b) Binary mask. (c) Guided matte from [40]. (d) Trimap.

4 Automatic Trimap Generation

In this section, we describe an automatic trimap generation process which utilizes depth and color information in two steps: *binary segmentation* and *unknown region detection*. These two steps are fully automatic without user inputs.

4.1 Motivation

Unlike single image matting, the EPI correspondences of a light field image implicitly encodes depth information. Following our assumption in Sec 3.3 that the foreground and background are located at different depth, we can estimate the trimap automatically. However, directly estimate the trimap is difficult because of the color mixing effects along the foreground object boundaries. Instead, we propose a two phases approach to first estimate salient foreground and background regions through binary segmentation by graph cut. Then, we detect the unknown regions by analyzing the local color distribution along object boundaries. If a local region contain mixed colors, it is likely that the local regions belongs to unknown regions. In contrast, if the color distributions are very distinctive, it is likely that the local region is from sharp edges. Based on this observation, we measure the KL divergence of local color distribution to evaluate the entropy of color mixing in order to find the optimal trimap.

4.2 Binary Segmentation

We use the proposed method in [39] to estimate a depth map from a light-field image. In practice, depth map estimations from light field images contain errors. These errors can be due to homogeneous regions, repeated textures, occlusion, and parallax. Also, aliasing, aberration, and inconsistent color tone in a light field image can further hinder the depth

map estimation. The inaccurate depth map can cause large errors in the binary segmentation.

To overcome the aforementioned problems, we use color information together with depth information for binary segmentation. Instead of using all pixels to compute the color distribution of foreground and background, we use only the reliable pixels whose depth values are within the closest and farthest 20% from the camera. To globally model color distribution from only selected reliable pixels, we adopt Gaussian mixutre model (GMM). In other words, we collect reliable pixels from an input image and an estimated depth map, then learn GMM parameters to represent global distributions of foreground and backgound. From the initial binary segmentation by depth map, we model color and depth distributions of foreground and background using the GMM:

$$p(i) = \sum_{k=1}^{N} \omega_k \cdot G(\pi(i), \mu_k, \Sigma_k), \tag{6}$$

where $p(i)$ and $\pi(i)$ are the probability and a query variable at pixel i. N, ω_k, μ_k, Σ_k are the number of the GMM components, a weight, an expectation and a covariance matrix for each kth component, respectively. p_F and p_B are computed using Eq. (6) using the estimated foreground and background GMM. A similar cost function for binary segmentation was also used in [41]. The GMM parameters are estimated by expectation maximization (EM) algorithm [42] for color and depth of foreground and background separately. The number of GMM components is decided by the number of peaks in the histogram of samples. To construct samples for GMM estimation, we consider pixels whose depth value is within the closest and farthest 20% as confidence regions for foreground, and background samples respectively. Using the estimated GMM parameters of highly confident foreground and background regions, we obtain an initial binary mask with confidence measures as follow:

$$\begin{aligned} \hat{b}(i) &= \begin{cases} 1 & \text{if } p_F(i) \geq p_B(i), \\ 0 & \text{ohterwise.} \end{cases} \\ m(i) &= \max\left(\frac{p_F(i)}{p_F(i)+p_B(i)}, \frac{p_B(i)}{p_F(i)+p_B(i)}\right), \end{aligned} \tag{7}$$

where $\hat{b}$, m, p_F, and p_B are the initial binary mask, the confidence measure, and probability density function of foreground and background respectively. From Eq. (7), the initial binary masks for color and depth ($\hat{b}_c, \hat{b}_d$) with its confidence maps ($\hat{m}_c, \hat{m}_d$) are computed. While a target foreground object is usually connected, Eq. (7) is per-pixel operation without any spatial constraint. In order to encode spatial coherences, the final binary segmentation is performed by solving following objective function:

$$\begin{aligned} b^* = \arg\min_b m_c \cdot \left\| b - \hat{b}_c \right\| + m_d \cdot \left\| b - \hat{b}_d \right\| \\ + \lambda \sum_i \sum_{j \in \mathcal{N}} \left\| b(i) - b(j) \right\|, \end{aligned} \tag{8}$$

where λ, and $\mathcal{N}$ are balancing weight factors and neighborhood for smoothness term. Eq. (8) can be solved effectively by the graph cut algorithm [43].

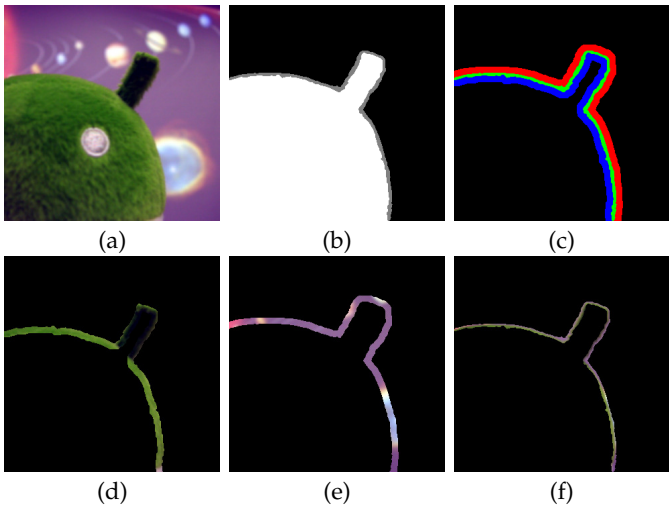


Fig. 4. (a) An image of central view. (b) An estimated trimap from Eq. (10). (c) Regions for clustering. Blue, green, and red areas are for foreground, unknown, background color samples. (d,e,f) show color samples for clustering. In this example, foreground and background color distributions are well separated, and unknown distribution has mixtures of foreground and background colors.

4.3 Unknown Region Detection

An optimal trimap is a trimap whose unknown regions cover only the regions where color mixing between foreground and background occurs. In order to estimate an optimal trimap, it is necessary to consider contexts of an image such as color and object structures. However, a trimap generated by simple morphological operations from a binary mask ignores contexts of an image by taking all neighboring pixels of boundaries as unknown regions. In order to encode contexts of an image, we first compute an initial alpha matte using the guided filter [40]. The estimated binary mask is filtered under the guidance of a RGB image to compute an initial alpha matte. After the guided filtering, boundary structures of a RGB image are transferred to the boundaries of the binary mask.

In practice, the result of the guided filter can vary significantly with different window size of the guided filter. To choose a proper window size w, we design a following criterion:

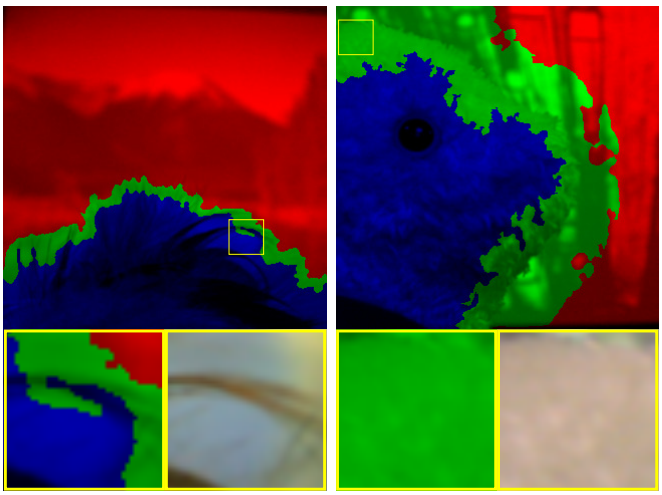
$$w^* = \arg\max_w \frac{KL(T_w)}{R(T_w)}, \tag{9}$$

where $R(T_w)$ and $KL(T_w)$ are ratio of unknown regions over entire pixels, and KL-divergence given a trimap T_w, respectively. A trimap T_w from an initial alpha matte for each window size of the guided filter w is defined by

$$T_w(i) = \begin{cases} 1 & \text{if } \hat{\alpha}_w(i) \geq 1 - \epsilon, \\ 0 & \text{if } \hat{\alpha}_w(i) \leq \epsilon, \\ 0.5 & \text{ohterwise,} \end{cases} \tag{10}$$

where $\hat{\alpha}_w$ is an estimated alpha matte given window size w. ϵ is 0.01 in our experiments. The KL-divergence of foreground and background color distributions is defined by

$$KL(T_w) = \sum_{j=1}^{N_U} \left\| U_{T_w}(j) - F_{T_w}(j') \right\| \log \frac{\left\| U_{T_w}(j) - F_{T_w}(j') \right\|}{\left\| U_{T_w}(j) - B_{T_w}(j'') \right\|}, \tag{11}$$



(a) (b)

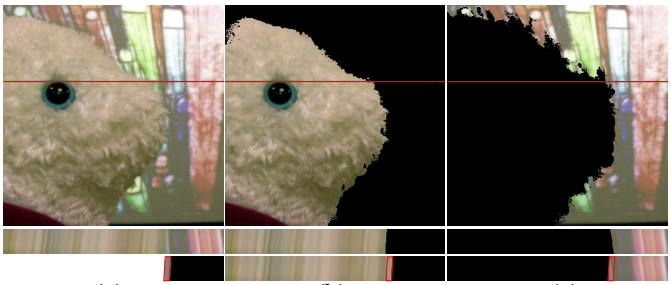
Fig. 5. The effects of window size, blue, green, and red areas are foreground, unknown, and background regions of a estimated trimap from Eq. (10) respectively. (a) Long hair structures with small window size. Note that background color (sky blue) are included in foreground regions. (b) Short hair structures with large window size. Note that unknown regions of a triamp are unnecessarily large. Window size of guided filter should be changed properly according to the object structures.

where U_{T_w}, F_{T_w}, and B_{T_w} are color cluster centers via a two level hierarchical clustering as introduced in [7]. Intuitively, by maximizing Eq. (11), we are searching the best window size which distance of color distributions of unknown regions to foreground, and background regions are maximized.

In order to model the local color distribution of foreground, background and unknown regions, we adopt a two-level clustering approach. A similar clustering approach was also used in the comprehensive sampling matting [7]. The first level clustering is performed on color space followed by the second level clustering on spatial domain. Each cluster has mean color and spatial position. N_U is the total number of color cluster centers of unknown regions given trimap T_w. j is a index of an unknown region cluster, $j^{'}$ and $j^{''}$ are indexes of foreground and background clusters closest to j in spatial domain, respectively. As illustrated in Fig. 4, U_{T_w} is estimated in unknown regions of trimap T_w. F_{T_w} and B_{T_w} are clustered in nearby unknown areas.

For each set of window sizes, we compute trimaps T_w using Eq. (10). From the computed trimaps, we solve Eq. (9) get the optimal window size w^* using brute-force optimization. We choose a trimap which is made from the best window size w^* of the guided filter among all sets of computed trimaps. As demonstrated in Fig. 5, if the window size is too small, area of unknown regions are too narrow to cover all overlapped areas especially for long hair structures. In contrast, if the window size is too large, unnecessary pixels are redundantly included. Therefore, it is important to decide proper window size for the guided filter.

Finally, the optimal trimap of the center view is propagated to other sub-images using the estimated foreground EPI. Fig. 6 shows EPI images masked out using a propagated trimap. The automatic trimap generation (implemented in Matlab) takes around $10 \sim 30$ seconds to process a light-field image with resolution 320×320 on a machine with Intel i7 CPU.



(a) (b) (c)

Fig. 6. (a) A center view image with EPI an image. (b,c) Foreground and background images masked using trimap with EPI images.

5 Consistent Matting for Light-Field Images

In this section, we present our algorithm for consistent light-field image matting. We describe how to utilize EPI constraint for consistent alpha mattes across the multiple sub-images of a light-field image. Using the EPI constraint, we derive two solutions to estimate alpha mattes when correspondence samples along EPI line are known, and unknown. To achieve further consistency, multi-image matting Laplacian with EPI smoothness constraints will be introduced. We assume a trimap of a light-field image is already computed as discussed in Sec. 4.

5.1 Color Sample Selection

In previous sampling based matting algorithms, color samples are selected to minimize the linear composite error defined by the matting equation in Eq. (1). A major challenge in this process is that there are multiple pairs of foreground and background samples that can minimize the error but the estimated alpha values can be totally different. Researches in sampling based matting algorithm have extensively focused on how to resolve this ambiguity by using different cues or making different assumptions about the true color samples. In this section, we describe how to resolve this ambiguity by using the EPI correspondence defined in the previous section.

In light-field image matting, when the foreground EPI intercept with the background EPI, we have two different cases as illustrated in Fig. 7.

Case 1: Background (foreground) samples along EPI of foreground (background) are known.

This case happens when a background sample is partially occluded in one sub-image, but is disoccluded in another sub-image. The disoccluded background samples can be easily detected using the background color sample correspondence defined by the background EPI.

To estimate the color of a foreground sample with known background colors, we can derive multiple equations along

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

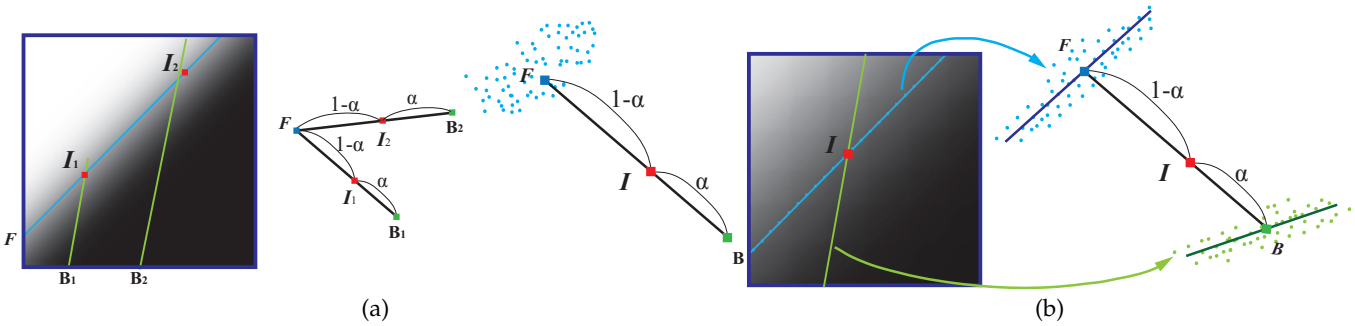


Fig. 7. (a) Case 1: If $B_1 \neq B_2$, we solve the alpha in a closed form. If $B_1 = B_2$, we solve the alpha using the comprehensive sample set. (b) Case 2: Foreground and Background samples are solved individually along its EPI, and the median F and B are selected to solve the alpha.

the foreground color sample correspondence defined by the foreground EPI:

$$\begin{aligned} I_1 &= \alpha F + (1-\alpha)B_1, \\ &\vdots \\ I_n &= \alpha F + (1-\alpha)B_n, \end{aligned} \tag{12}$$

where $\{I_1, \ldots, I_n\}$ are the observed intensity along the foreground color sample correspondence, and $\{B_1, \ldots, B_n\}$ are the known background colors. Thus, we have number of equations more than or equal to the number of unknown when $n \geq 2$. When $n = 2$, we can obtain the alpha in a closed form:

$$\hat{\alpha} = 1 - \frac{I_1 - I_2}{B_1 - B_2}. \tag{13}$$

When $n > 2$, we solve the alpha using the least square error method by computing the weighted average α across the solution of all pairs of pixels in the foreground color sample correspondences with known background color. The weighting factors for each α are determined by distance between two background color samples. Longer distance has larger weight coefficient because it is more reliable to estimate α value than the inverse case where the denominator is close to zero. With the estimated alpha, $\hat{\alpha}$ and the known background colors, B_i, the foreground color, F, can be computed accordingly.

When we estimate the color of background sample with known foreground EPI line correspondences, Eq. (14), Eq. (15) can be modified as:

$$\begin{aligned} I_1 &= \alpha F_1 + (1-\alpha)B, \\ &\vdots \\ I_n &= \alpha F_n + (1-\alpha)B, \end{aligned} \tag{14}$$

$$\hat{\alpha} = \frac{I_1 - I_2}{F_1 - F_2}. \tag{15}$$

The same strategies are used to get final $\hat{\alpha}$ and B with known the foreground color F_i along the EPI.

In a degenerated case when color samples along the EPI line are identical, e.g. homogeneous color, we use the comprehensive sample set collected from the known region to estimate the alpha. In order to efficiently process color samples from the known regions, we follow the steps in [7] to construct a comprehensive color sample sets of foreground and background in a light-field image. Similar to color clustering for KL-divergence computation in Sec. 4.3, we use the two-level hierarchical clustering process. This provides us the comprehensive sample set which covering all possible foreground and background colors in a light-field image.

To select an foreground sample from collected sample set given an background sample, or to select an background sample from collected sample set given an foreground sample, we use brute-force optimization based on the following objective function:

$$O(F,B) = C(F,B)^{e_C} \times S(F,B)^{e_S} \times Z(F,B)^{e_Z}, \tag{16}$$

where $C(F,B)$, $S(F,B)$ and $Z(F,B)$ measure chromatic distortion, spatial distance and depth compatibility of a pair of (F,B). e_C, e_S and e_Z are weighting factors for balancing each term. Note that because one of F and B is known, therefore the computational cost for the optimization is not heavy.

In our optimization, $C(F,B)$ is commonly used in color sampling based matting methods to fit a pair of (F,B) to matting equation Eq. (1) and is defined by

$$C(F,B) = \exp(-\|I - (\hat{\alpha}F + (1-\hat{\alpha}B))\|), \tag{17}$$

where $\hat{\alpha}$ is estimated alpha mattes from Eq. (2) given a pair of (F,B). Since one of F and B is already known in the degenerated case, Eq. (17) play a role in enforcing line constraint that the I, F and B must be located along the same line in 3D color space. $S(F,B)$ represents the spatial distance between a selected color sample and a query pixel as following:

$$S(F,B) = \exp(-\|i - i_F\|) \times \exp(-\|i - i_B\|), \tag{18}$$

where i, i_F and i_B are spatial coordinate of a pixel, center of foreground and background color clusters respectively. Since one of F and B is known, only one of the left or right term in Eq. (18) is needed. $Z(F,B)$ is a term about depth compatibility of an estimated $\hat{\alpha}$ from Eq. (2) with depth values of foreground and background:

$$Z(F,B) = \hat{\alpha} \times \frac{\|z - z_B\|}{N_z} + (1-\hat{\alpha}) \times \frac{\|z - z_F\|}{N_z}, \tag{19}$$

where $N_z = \|z - z_F\| + \|z - z_B\|$, z, z_F, and z_B are a depth of a pixel, foreground and background depth values respectively. When both of $\hat{\alpha}$ and z are close to foreground or background, Eq. (19) obtains a high response.

By maximizing Eq. (16), an optimal foreground (background) color from the comprehensive color sample sets is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

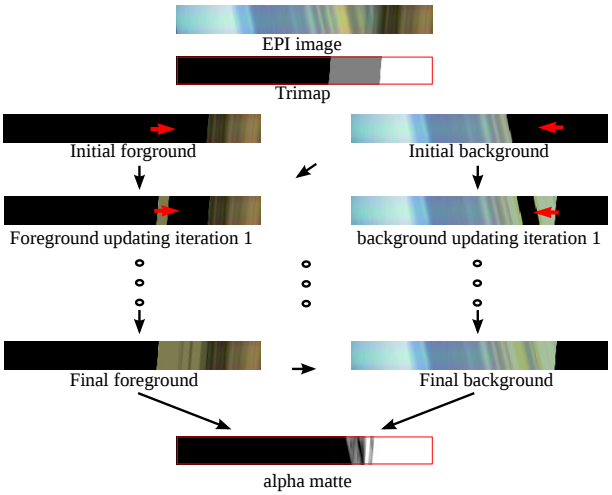


Fig. 8. Illustration of color sample selection via propagation along with EPI line. Red arrows show directions of updating. This process reduces the Case 2 to the Case1.

obtained with corresponding $\hat{\alpha}$ given a background (foreground) color.

Case 2: Background (foreground) samples along EPI of foreground (background) are not known.

When alpha matte area is large, background (foreground) pixels closed to foreground (background) region will be occluded/partially occluded in all sub-images. In this case, we apply the following method to estimate foreground and background samples. Again, we assume the foreground EPI and the background EPI are misaligned.

For a pixel with different foreground and background color sample correspondence, we first compute the foreground and background sample pairs for each pixels along the foreground EPI and the background EPI independently. This process is done by maximizing Eq. (16) to select the optimal color samples from the comprehensive sample set collected from the known foreground and background regions. Unlike a degenerated case of Case 1, both of F and B are unknown, thus both terms in Eq. (18) are needed. Next, assuming that majority of the estimated color samples are correct, we apply a simple linear regression to fit a line to the estimated foreground and background samples. Note that when computing the foreground color, only the estimated foreground samples along the foreground EPI were used. This is the same case for the background color estimation. After fitting the color line, we sort the color samples along the estimated color line and choose the median foreground color and the median background color as the true foreground color and the true background color respectively. Once the true foreground and background colors are estimated, the alpha value can be computed from Eq. (2).

While this method is simple, we find that the method robust and reliable. In a degenerated case where all foreground and background colors along the two different EPIs are identical, this reduces the problem to the conventional setting of color sample estimation since the two EPIs do not provide additional information to assist the color sample selection.

Implementation: solving alpha in the order to reduce ambiguity.

Since pixels are interconnected by the foreground and the background EPI, and each pixel in the unknown region have two different set of correspondences, once the foreground color of a pixel with known background samples (Case 1) is estimated, the estimated foreground color can be used to estimate background color of a pixel where its background samples are unknown (Case 2). Using this strategy, we can significantly reduces the ambiguity in color sample selection by reducing the Case 2 scenario to the Case 1 scenario using the estimated color as the known color samples. Fig. 8 shows how foreground and background color samples are updated along with EPI line in alternating manner. In other words, alpha value of pixels are solved progressively from the boundary of foreground and background regions towards the center of unknown region.

5.2 Consistent Matting with the EPI smoothness term

The previous color sampling step estimates alpha value of each pixel independently although the selected color samples is guided by the EPI constraint. In this section, we describe the process to further improve the alpha matte by considering smoothness among neighboring pixels. This is also a common post-processing step in many previous matting algorithms.

Using the results $\hat{\alpha}$ from the Eq. (15)/Eq. (2) as a data term, and the smoothness term defined by the matting Laplacian matrix L [9], we can obtain the final alpha by:

$$\alpha = \arg\min \alpha^T L\alpha + \lambda(\alpha - \hat{\alpha})^T D(\alpha - \hat{\alpha}) \qquad (20)$$

where λ is a weighting parameter, and D is a diagonal matrix. Its diagonal element is a large constant for the known pixel, and a confidence $c = \exp(||I - (\hat{\alpha}F + (1-\hat{\alpha})B||^2/\sigma^2)$ for the unknown pixel.

In order to further consider the smoothness constraint along the EPI of the extracted foreground, we extend Eq. (20) to include an additional nonlocal smoothness term in L, and solve the alpha matte of multiple sub-images simultaneously. In particular, we extend Eq. (20) as follows:

$$\alpha = \arg\min \alpha^T \mathbf{L}\alpha + \lambda(\alpha - \hat{\alpha})^T \mathbf{D}(\alpha - \hat{\alpha}), \qquad (21)$$

where $\alpha = [\alpha_1, \alpha_2, \alpha_3]^\top$, and

$$\mathbf{L} = \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \qquad (22)$$

where L_{ii}, $i = \{1, 2, 3\}$, are the matting Laplacian of the sub-image I_i, and L_{ij}, $i \neq j$, are the cross sub-images smoothness term with each entry defined as:

$$L_{ij}(x, x') = \exp(-\frac{||I_i(x) - I_j(x')||^2}{2\sigma_c^2}),$$

if x and x' are the foreground EPI correspondence between I_i and I_j, and $L_{ij}(x, x') = 0$ if otherwise. For a sub-image I_1, I_2 is the sub-image next to I_1 in horizontal direction, and I_3 is the sub-image next to I_1 in vertical direction. Thus, we can solve the alpha matte of three sub-images simultaneously with consideration of the EPI smoothness of alpha matte across the sub-images. Although we can solve the alpha matte of sub-images altogether by further extending

Eq. (22) to include more sub-images, the computation cost increases dramatically deal to the large linear system. In experiments, we found that using more adjacent sub-images does not improve much in accuracy. Thus, we only solve the alpha matte of three sub-images simultaneously.

6 Experimental Results

6.1 Light-field matting dataset

We follow the steps in [15] to create a new dataset to evaluate the performance of matting algorithms applied on light-field images. We utilize two types of light-field cameras: Lytro [44] and lab-made light-field camera developed by [33]. The lab-made light-field camera produces larger disparities than Lytro camera. The method in [27] is used to decode the captured light-field images from its RAW image data. In order to derive a high-quality ground truth alpha matte, we placed the matting objects in front of a monitor, and we displayed four different single-colored background (i.e. black, red, green, blue). This gives us 45 sub-images (7×7 without four corners) of a light-field image. With the different monochrome color background, we apply the blue screen matting [45] to get the ground truth alpha matte by triangulation. To capture the images for testing, we change the background on the monitor with natural background images. Finally, the images were cropped at a bounding box that was casually drawn around the foreground objects, resulting in the test scenes.

Our dataset [1] consists of 18 testing images, and the foreground objects were chosen to cover different properties with hard and soft boundaries as well as transparent objects. Fig. 9 shows some of our testing images, trimaps and the ground truth alpha mattes. Trimaps in the second row are given by users and trimaps in the third row are generated by the methods in Sec. 4. The center view trimap is propagated to other sub-images automatically using the foreground EPI.

6.2 Evaluations

We evaluate the performance of our algorithm in term of RMSE and consistency. The RMSE is computed as follows:

$$RMSE(\alpha) = \sqrt{\frac{1}{N}\sum_i (\alpha_i^* - \alpha_i)^2}, \quad (23)$$

where α^* is the ground truth alpha and N is the total number of pixels. The consistency is evaluated as follows:

$$CONS(\alpha) = \sqrt{\frac{1}{N}\sum_i (\frac{1}{N_{EPI_i}} \sum_{j \in EPI_i} ||\alpha_i - \alpha_j||^2)}, \quad (24)$$

where α_i and α_j are the EPI correspondences defined by the foreground EPI in $x-s$ plane and $y-t$ plane respectively.

6.3 Comparisons

We compare the performance of our algorithm with the state-of-the-art matting algorithms: closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], and video matting [46]. Our tests are performed on two sets of trimap: user inputs and automatically generated trimaps as described in Sec. 4.

Tab. 1 summarizes the quantitative comparisons in terms of RMSE and consistency using user inputs and automatically generated trimaps, respectively. Regardless of trimap type, our algorithm achieves the best consistent across different sub-images and produces results with less RMSE than other algorithms. Generally, automatically generated trimaps produce similar results with user inputs. When foreground and background color distributions are highly overlapped around boundaries, regions of an automatically generated trimap will be broaden such as Data04, Data05, and Data07. In this cases, RMSE of automatically generated trimaps is higher than user inputs. However, as shown in Fig. 9, it requires much lobar and time to make a proper trimap while our automatic trimap generation takes only a few seconds.

For two types of trimaps, Fig. 10, Fig. 11, Fig. 12 and Fig. 13 show that our results qualitatively outperform previous matting methods on real light-field images taken from Lytro and lab-made light-field camera.

We also check consistency of an estimated alpha matte using refocusing. If the estimated alpha mattes are inconsistent, outliers will appear during the refocusing. We refocus light-field alpha mattes on an object and background plane, and combined the refocused objects with green background. As shown in Fig. 14, our results are more visually pleasing compared to results from previous methods. This is because our algorithm produces more consistent alpha mattes across the sub-images.

Since the trimap generation relies on depth information, we perform additional experiments to analyze effects of depth accuracy to the quality of trimap, and our final matting results. To obtain a depth map from a light-field image, we use two methods proposed by Tao *et al.* [39], and Jeon *et al.* [33] as shown in Fig. 15 (b)-(c). In addition, we use a Kinect-2 to capture depth information from the RGBD sensor. Since only one RGB image in the captured RGBD data, we placed a light field camera to capture a light field image at the same location as the RGB image from Kinect-2. The captured depth map from the Kinect-2 is aligned with the center view of a light-field image as shown in Fig. 15 (d). The depth quality from a RGBD sensor is usually better than the depth from light-field images. However, there is small misalignment error because the depth map and the light field image are captured from two different devices. Also, holes in the RGBD depth map near the object boundaries should be properly handled. For trimap generation, we only use the pixels whose depth value is within the closest and farthest 20% as confidence regions for foreground and background. However, since our trimap generation contains unknown region detection, and the estimated alpha mattes are forced to be consistent, inaccuracies along depth map boundary do not harm too much to our final results. Fig. 15 shows the comparisons where the depth maps are obtained using different methods, and the estimated alpha mattes are quite similar.

7 Limitation and Discussion

In this paper, we assume foreground and background have sufficient distance such that the directions of foreground EPI

1. https://sites.google.com/site/lightfieldmatting/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

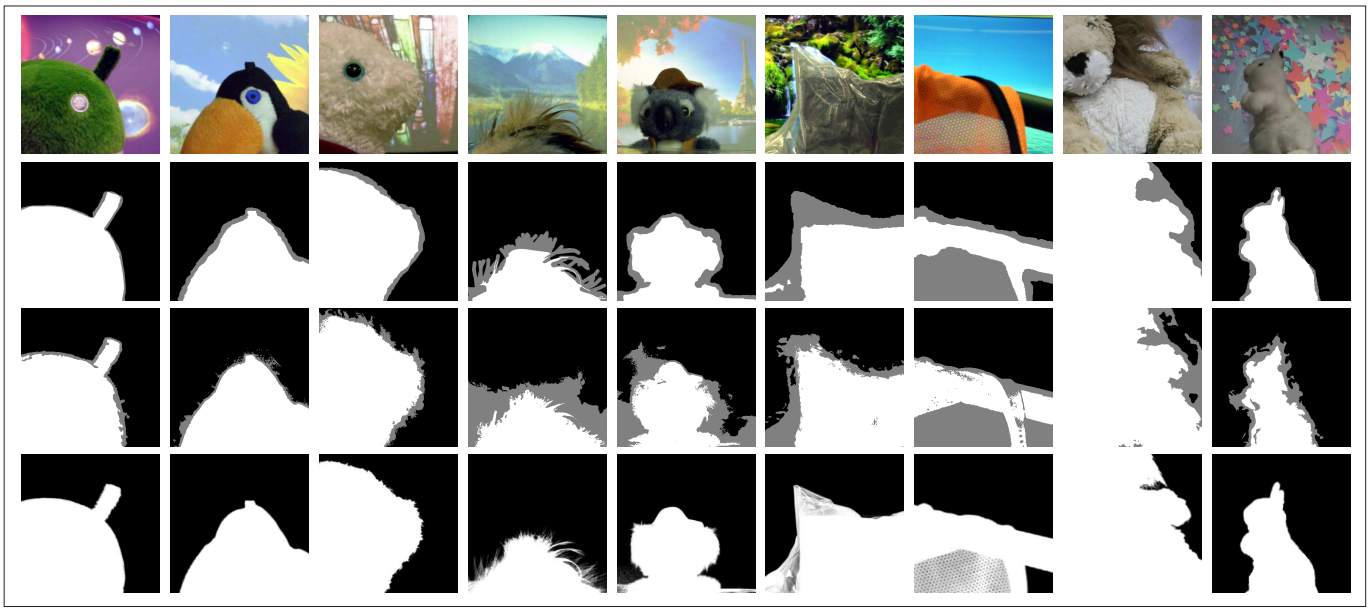


Fig. 9. Our dataset for testing. From top to bottom rows show center view images, trimaps given by users, estimated trimaps using a proposed method in Sec. 4 and ground truth alpha mattes respectively.

	Data01		Data02		Data03		Data04		Data05		Data06		Data07		Data08		Data09	
	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS	RMSE	CONS
Closed (user)	5.3	5.8	6.3	7.7	11.9	6.3	21.8	7.7	17.3	6.2	25.3	5.6	20	7.9	17.2	4.9	8.5	7.5
Closed (automatic)	5.9	5.7	5.2	7.2	15.7	6.3	59.6	4.9	40.4	6.1	28.1	6.2	28.5	6.8	22.3	4.7	20.8	6.8
NLM (user)	7.9	12.8	5.4	13.6	16.3	12.6	26.5	15.6	14.8	15.5	45.7	15.6	23.9	11.5	25.5	7.0	10.6	12.0
NLM (automatic)	9.1	12.4	9.0	12.5	20.2	12.1	64.8	9.9	41.9	39.0	14.3	16.1	28.1	8.7	32.7	7.6	26.1	10.7
KNN (user)	7.7	7.4	7.0	10.7	14.3	8.6	28.3	12.2	18.9	10.3	32.1	9.8	16.2	10.7	19.6	8.5	15.9	8.8
KNN (automatic)	8.4	7.8	10.8	10.5	18.2	9.4	50.0	11.8	31.9	11.0	28.4	10.9	26.1	10.3	43.7	8.3	29.0	9.2
COM (user)	5	5.9	6.6	8.3	11.5	6.7	25.9	9.9	17.7	6.8	22.1	7.1	17.6	11.7	19.2	7.9	8.3	8.5
COM (automatic)	4.2	6.0	5.2	7.4	14.8	6.6	48.2	7.9	28.8	7.8	28.5	7.1	16.7	11.6	26.4	8.0	17.6	8.6
VIDEO (user)	5.6	6.3	7.2	8.3	12.8	6.2	21.9	8.1	13.1	6.7	23.7	5.3	19.8	8.2	19.8	4.9	8.3	8.7
VIDEO (automatic)	4.8	6.1	5.5	7.9	15.1	6.4	54.4	4.3	23.2	6.2	27.7	6.1	27.4	7	24.4	5.3	17.9	7.6
OUR (user)	5	2.6	5.9	4.1	10.4	3.6	21.2	5.0	13.1	4.0	17.2	4.2	10.8	3.9	14.6	3.1	6.7	3.6
OUR (automatic)	4.2	2.5	5.0	4.0	13.7	3.9	36.3	3.7	28.0	4.1	24.7	5.1	10.1	3.8	16.5	3.5	15.3	4.1

TABLE 1
Quantitative Comparisons in term of RMSE and Consistency with trimaps given by user and automatically generated trimaps.

and background EPI are very different from each other. If this assumption is violated, it is difficult to utilize light-field properties and then our approach is not much effective. To satisfy our assumption, it is better to place the foreground object closed to the lightfield camera and far from the background, which results in very distinctive EPI gradients between foreground and background.

Due to the smaller directional resolution of light-field images, computing foreground, background samples and alpha value is sensitive to noise. Therefore we only apply case 1 method when the number of disoccluded pixel is more than 4 along the EPI, and it provides certain robustness against image noise. To make more robust to noise, some appropriate weighting factor can be adopted by estimating confidence level of each pixels on EPI line.

8 Conclusion

In this paper, we have presented automatic trimap generation and consistent matting for a light-field image. Our proposed method is fully automatic and does not require user-assistances.

In the automatic trimap generation, a rough binary mask is computed using depth and color information, then unknown regions are detected by analyzing color distribution along the boundaries of binary mask. To analyze and make a trimap, we use guided filter and KL-divergence measure. Using the generated trimap, we estimates consistent alpha mattes of a foreground object across multiple sub-image in a light-field image. By using the EPI constraint, we can define different sets of pixel correspondences for foreground and background. In the color sample selection, we have presented a method to estimate foreground samples alpha with known and unknown background samples. In addition, we have introduced a method to include the EPI smoothness constraint and proposed to solve alpha matte of multiple sub-images simultaneously.

In the experimental evaluations, we have created a new dataset with ground truth alpha mattes using two types of light-field cameras to quantitatively compare the performance of our algorithm with the performance of state-of-the-art image matting algorithms. Our algorithm outperforms previous work in term of both RMSE and consistency. As for future work, we are interested in extending our work in other applications that utilize light-field image data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

Fig. 10. Qualitative comparisons on Lytro dataset with user inputs. We compare our estimated alpha mattes with results from previous methods. Top to bottom: Input images, results from closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], video matting [46], our method, and ground truth.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680).

REFERENCES

[1] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *IEEE CVPR*, 2001.
[2] J. Wang and M. F. Cohen, "Optimized color sampling for robust matting," in *IEEE CVPR*, 2007.
[3] C. Rhemann, C. Rother, and M. Gelautz, "Improving color modeling for alpha matting," in *British Machine Vision Conference (BMVC)*, 2008.
[4] E. S. L. Gastal and M. M. Oliveira, "Shared sampling for real-time alpha matting," in *Eurographics*, 2010.
[5] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A global sampling method for alpha matting," in *IEEE CVPR*, 2011.
[6] E. Shahrian and D. Rajan, "Weighted color and texture sample selection for image matting," in *IEEE CVPR*, 2012.
[7] E. Shahrian, D. Rajan, B. Price, and S. Cohen, "Improving image matting using comprehensive sampling sets," in *IEEE CVPR*, 2013.
[8] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum, "Poisson matting," *ACM TOG*, vol. 23, no. 3, pp. 315–321, 2004.
[9] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. on PAMI*, vol. 30, no. 2, pp. 0162–8828, 2008.
[10] Y. Zheng and C. Kambhamettu, "Learning based digital matting," in *IEEE ICCV*, 2009.
[11] K. He, J. Sun, and X. Tang, "Fast matting using large kernel matting laplacian matrices," in *IEEE CVPR*, 2010.
[12] P. Lee and Y. Wu, "Nonlocal matting," in *IEEE CVPR*, 2011.
[13] Q. Chen, D. Li, and C.-K. Tang, "Knn matting," in *IEEE CVPR*, 2012.
[14] X. Chen, D. Zou, S. Z. Zhou, Q. Zhao, and P. Tan, "Image matting with local and nonlocal smooth priors." in *IEEE CVPR*, 2013.
[15] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, "A perceptually motivated online benchmark for image matting," in *IEEE CVPR*, 2009.
[16] D. Cho, S. Kim, and Y.-W. Tai, "Consistent matting for light field images," in *IEEE ECCV*, 2014.
[17] D. Cho, Y.-W. Tai, and I. S. Kweon, "Natural image matting using deep convolutional neural networks," in *IEEE ECCV*, 2016.
[18] O. Wang, J. Finger, Q. Yang, J. Davis, and R. Yang, "Automatic natural video matting with depth," in *Pacific Conference on Computer Graphics and Applications*, 2007.
[19] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand., "Defocus video matting," in *ACM SIGGRAPH*, 2005.
[20] N. Joshi, W. Matusik, and S. Avidan, "Natural video matting using camera arrays," in *ACM SIGGRAPH*, 2006.
[21] C. Rhemann, C. Rother, A. Rav-Acha, and T. Sharp., "High resolution matting via interactive trimap segmentation," in *IEEE CVPR*, 2008.
[22] C. Z. Bei He, Guijin Wang, "Iterative transductive learning for automatic image segmentation and matting with rgb-d data,"

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

Fig. 11. Qualitative comparisons on Lytro dataset with automatically generated trimaps. We compare our estimated alpha mattes with results from previous methods. Top to bottom: Input images, results from closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], video matting [46], our method, and ground truth.

Journal of Visual Communication and Image Representation, vol. 25, no. 5, pp. 1031–1043, 2014.

[23] S.-H. Kim, Y.-W. Tai, J. Park, and I. S. Kweon, "Multi-view object extraction with fractional boundaries," *IEEE TIP*, vol. 25, no. 8, pp. 3639–3654, 2016.

[24] J. Sun, Y. Li, S. Bing, and K. H. yeung Shum, "Flash matting," in *ACM SIGGRAPH*, 2006.

[25] S. Kim, Y.-W. Tai, Y. Bok, H. Kim, and I. S. Kweon, "Two-phase approach for multi-view object extraction," in *ICIP*, 2011.

[26] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Tech. Rep., 2005.

[27] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *IEEE ICCV*, 2013.

[28] D. G. Dansereau and L. T. Bruton, "Gradient-based depth estimation from 4d light fields," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2004.

[29] T. E. Bishop, S. Zanetti, and P. Favaro, "Plenoptic depth estimation from multiple aliased views," in *IEEE ICCV Workshops*, 2009.

[30] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D lightfields," in *IEEE CVPR*, 2012.

[31] S. Wanner, C. Straehle, and B. Goldluecke, "Globally consistent multi-label assignment on the ray space of 4d light fields," in *IEEE CVPR*, 2013.

[32] B. Goldluecke and S. Wanner, "The variational structure of disparity and regularization of 4d light fields," in *IEEE CVPR*, 2013.

[33] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *IEEE CVPR*, 2015.

[34] S. Wanner and B. Goldluecke, "Reconstructing reflective and transparent surfaces from epipolar plane images," 2013.

[35] J. Fiss, B. Curless, and R. Szeliski, "Light field layer matting," in *IEEE CVPR*, 2015.

[36] Lytro, "The lytro camera," https://www.lytro.com.

[37] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *IJCV*, vol. 1, no. 1, pp. 7–55, 1987.

[38] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding (CVIU)*, vol. 97, no. 1, pp. 51–85, 2005.

[39] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE ICCV*, 2013.

[40] K. He, J. Sun, and X. Tang, "Guided image filtering," in *IEEE ECCV*, 2010.

[41] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM TOG*, vol. 23, no. 3, pp. 303–308, 2004.

[42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[43] R. Z. Vladimir Kolmogorov, "Multi-camera scene reconstruction via graph cuts," in *IEEE ECCV*, 2002.

[44] Lytro, "The lytro camera," https://www.lytro.com.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

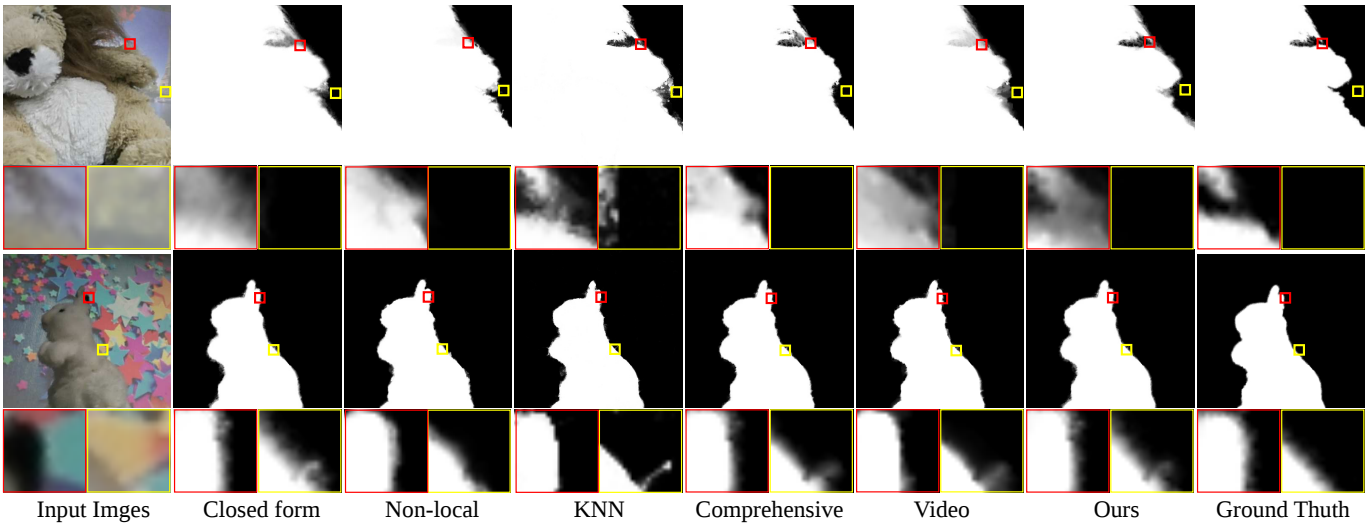


Fig. 12. Qualitative comparisons on Samsung lihgt field camera dataset with user inputs. We compare our estimated alpha mattes with results from previous methods. Left to right: Input images, results from closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], video matting [46], our method, and ground truth.

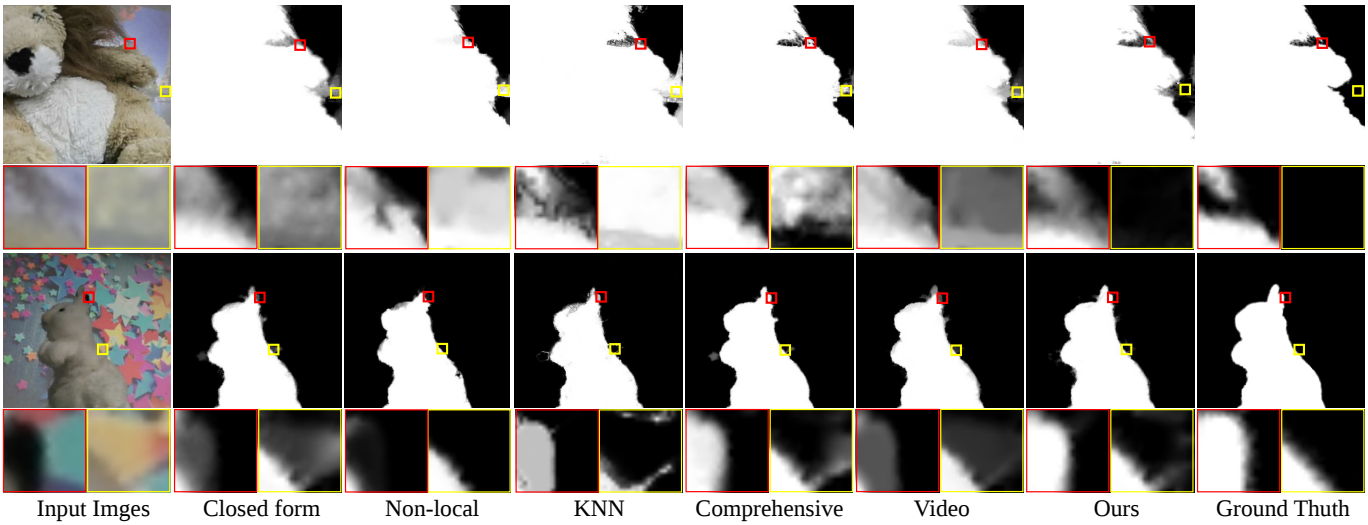
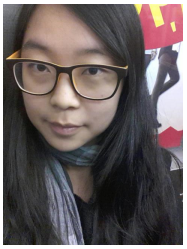


Fig. 13. Qualitative comparisons on Samsung lihgt field camera dataset with automatically generated trimaps. We compare our estimated alpha mattes with results from previous methods. Left to right: Input images, results from closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], video matting [46], our method, and ground truth.

[45] A. R. Smith and J. F. Blinn, "Blue screen matting," in *ACM SIGGRAPH*, 1996.
[46] I. Choi, M. Lee, and Y.-W. Tai, "Video matting using multi-frame nonlocal matting laplacian," in *IEEE ECCV*, 2012.

Donghyeon Cho is a Ph.D. student in EE department of KAIST, South Korea. From Sept 2014 to Feb 2015, he worked as a full-time student internship in the Microsoft Research Asia (MSRA). He received the MS degree in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea in 2014. He received the BS degree in information and communication engineering from Inha University, Incheon, Korea, in 2012. His research interests include image/video processing, computer vision and machine learning. He is a student member of the IEEE.



Sunyeong Kim is a software engineer at Perples, Inc. She received the BS and MS degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejon, Korea in 2010 and 2012 respectively. Her research interests include computer vision and image processing, especially in computational photography.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2016.2606397, IEEE Transactions on Pattern Analysis and Machine Intelligence

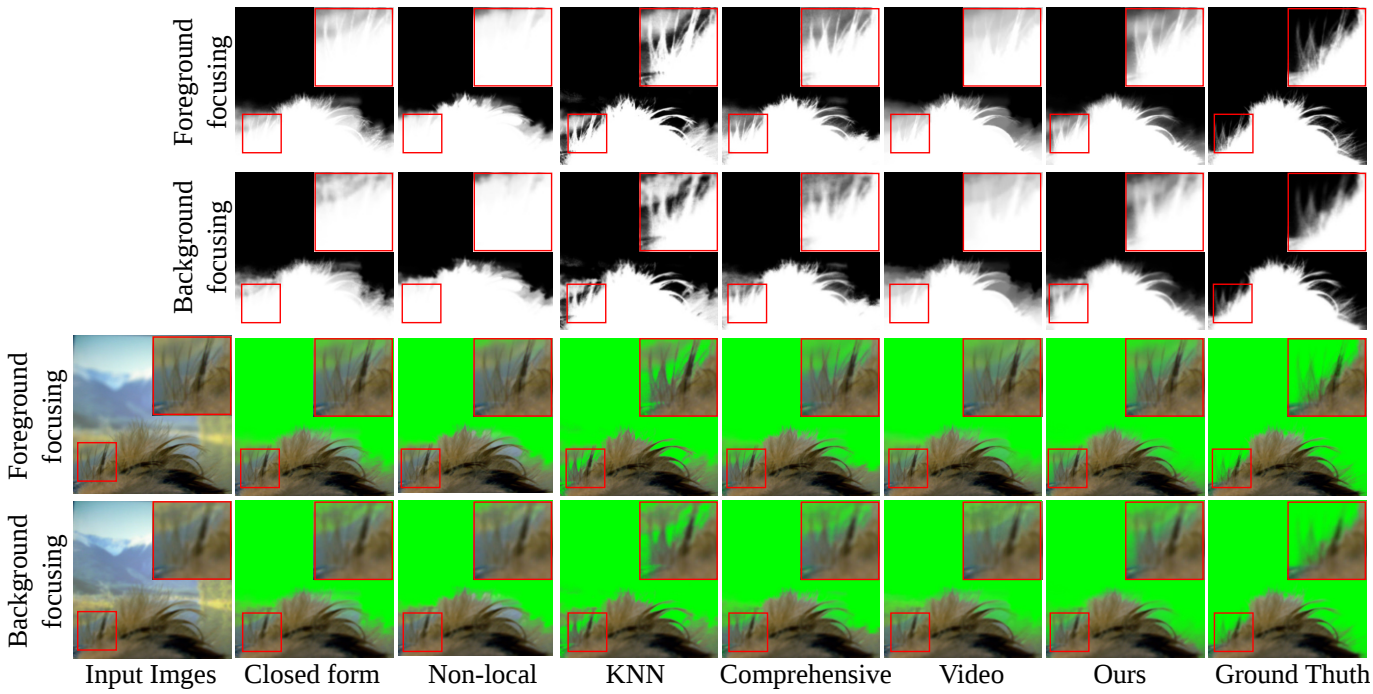


Fig. 14. Refocused alpha mattes for consistency checking. Left to right: Input images, results from closed form matting [9], non-local matting [12], KNN matting [13], comprehensive matting [7], video matting [46], our method, and ground truth. The first two rows are estimated alpha mattes and the last two rows are the combined image with the green background. Images in the first and third rows are focused on an object, and images in the second and fourth rows are focused on background.

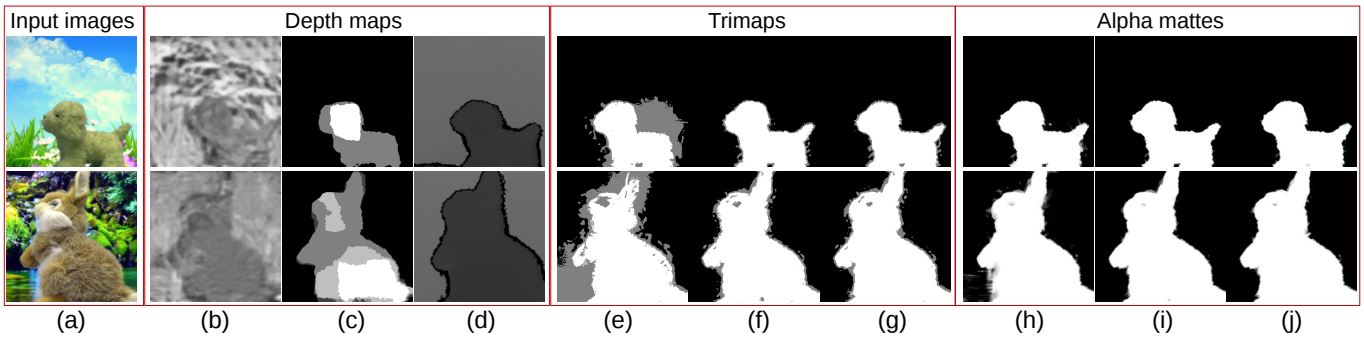


Fig. 15. Effects of depth accuracy to the trimap generation. (a) Input center view images. (b)-(d) Depth maps from [39], [33] and Kinect-2. (e)-(g) Estimated trimaps using (b)-(d). (h)-(j) Estimated alpha mattes using (e)-(g).

Yu-Wing Tai is a principle research scientist of SenseTime Group Limited. He was an associate professor working at the Korea Advanced Institute of Science and Technology (KAIST) from July 2009 to August 2015. He received his Ph.D. degree in Computer Science from the National University of Singapore (NUS) in 2009. From Sept 2007 to June 2008, he worked as a full-time student internship in the Microsoft Research Asia (MSRA). He was awarded the Microsoft Research Asia Fellowship in 2007, and the KAIST 40th Anniversary Academic Award for Excellent Professor in 2011 respectively. He received a M.Phil and B.Eng (First Class Honors) degree in Computer Science from the Hong Kong University of Science and Technology (HKUST) in 2005 and 2003 respectively. He is a senior member of the IEEE.

In So Kweon received the B.S. and M.S. degrees in Mechanical Design and Production Engineering from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively, and the Ph.D. degree in Robotics from the Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, in 1990. He worked for the Toshiba R&D Center, Japan, and joined the Department of Automation and Design Engineering, KAIST, Seoul, Korea, in 1992, where he is now a professor with the Department of Electrical Engineering. His research interests are sensor fusion, color modeling and analysis, visual tracking, and visual SLAM. He was the general chair for the Asian Conference on Computer Vision 2012 and he is on the honorary board of the International Journal of Computer Vision (IJCV). He has been serving as a director for the Personal Plug and Play DigiCar Center which is one of the National Core Research Center since 2010. He was a member of '*Team KAIST*' which won the first place in DARPA Robotics Challenge Finals 2015.