Dual-focus stereo imaging

Feng Li
University of Delaware
Computer and Information Sciences
101 Smith Hall, 18 Amstel Avenue
Newark, Delaware 19716
E-mail: feli@cis.udel.edu

Jian Sun
Microsoft Research Asia
5/F, Beijing Sigma Center
No. 49, Zhichun Road, Hai Dian District
Beijing, China 100190

Jue Wang
Adobe Systems
801 North 34th Street
Seattle, WA 98103-8882

Jingyi Yu
University of Delaware
Computer and Information Sciences
Newark, Delaware 19716

Abstract. *We present a novel stereo imaging technique called dual-focus stereo imaging or DFSI. DFSI uses a pair of images captured from different viewpoints and at different foci, but with identical wide aperture size. Each image in an DFSI pair exhibits different defocus blur, and the two images form a defocused stereo pair. To model defocus blur, we introduce a defocus kernel map (DKM) that computes the size of the blur disk at each pixel. We derive a novel* disparity defocus constraint *for computing the DKM in DFSI, and integrate DKM estimation with disparity map estimation to simultaneously recover both maps. We show that the recovered DKMs provide useful guidance for segmenting the in-focus regions. We demonstrate using DFSI in a variety of imaging applications, including low-light imaging, automatic defocus matting, and multifocus photomontage.*
 [DOI: 10.1117/1.3500802]

1 Introduction

Recent advances in digital imaging and computer vision[1] have led to the development of many new imaging systems. The use of specially designed apertures,[2–5] lenses,[6] flashes,[7] or shutters[8] has become a common practice for various imaging applications. For example, coded[4] or color-filtered apertures[2] can be used to robustly recover scene depth from a single shot. However, these imaging systems require delicately modifying the camera, and thus are not easily accessible for consumers.

In this work, we propose a novel stereo imaging technique that we call dual-focus stereo imaging (DFSI), which only uses off-the-shelf cameras. DFSI captures a pair of images from different viewpoints and at different scene foci, but with identical wide aperture size, as shown in Fig. 1. Wide apertures allow more light to be admitted to the camera and are suitable for low-light and fast motion imaging. However, they also lead to shallow depth of field (DOF), where pixels are sharp around what the lens is focusing on and blurred elsewhere. Hence, each image in DFSI exhibits different defocus blur and the two images form a defocus stereo pair, as shown in Fig. 2.

We first present a new framework to uniformly model defocus blur and parallax in a DFSI pair using a defocus kernel map (DKM). A DKM computes the size of the blur disk at each pixel via the *defocus constraint*. We derive a novel *disparity defocus constraint* for DFSI pairs and integrate DKM estimations with the graph-cut-based disparity map estimation to simultaneously recover both maps. The recovered DKMs provide useful guidance for segmenting the in-focus regions in each image.

We demonstrate a broad class of imaging applications using the DFSI technique. We first apply DFSI for enhancing low-light imaging by coupling wide apertures with fast shutters. We segment and warp the in-focus region using the recovered DKMs and disparity map. DFSI generates images

Paper 10021R received Feb. 9, 2010; revised manuscript received Jul. 26, 2010; accepted for publication Aug. 30, 2010; published online Dec. 6, 2010

1017-9909/2010/19(4)/043009/12/$25.00 © 2010 SPIE and IS&T.

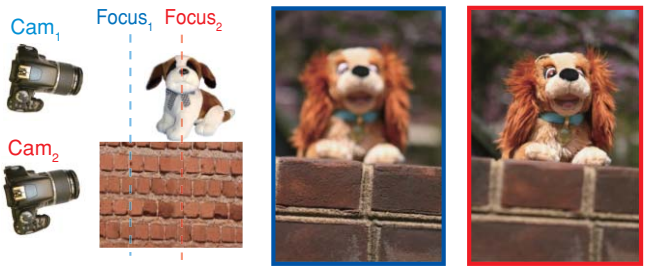


Fig. 1 A dual-focus stereo (DFS) pair.

with high sharpness, low noise level, and minimal motion blur in low light. We then use DFSI for automatic defocus matting. While existing defocus matting systems[9] require images to be captured at or warped to the same viewpoint, DFSI not only allows but also efficiently uses parallax between images. Finally, we demonstrate synthesizing a multifocus photomontage from a DFSI pair. To maintain smooth transitions between the warped and captured regions, we develop a novel adaptive blurring technique using the recovered DKMs.

The primary contributions of this work are:

- a dual-focus stereo imaging (DFSI) technique that uses a pair of images captured from different viewpoints and at different scene foci
- a novel *disparity defocus constraint* that uniformly models the defocus kernel maps (DKMs) and the disparity map in a DFSI pair
- a robust algorithm based on the *disparity defocus constraint* for recovering the DKMs and the disparity map
- a class of DFSI-based algorithms for various imaging applications, including low-light imaging, automatic defocus matting, and multifocus photomontage.

2 Related Work

In this section, we briefly review several research areas that are closely related to this work.

2.1 *Depth from Defocusing*

Recovering scene depth from multiple defocused images is a well-explored problem in computer vision. Most existing approaches require capturing a large number of images from the same viewpoint with different focuses.[10,11] By minimizing the blur, scene depth can then be estimated. For example, Ref. 12 captures all possible combinations of aperture and focus settings to recover very high quality depth maps. It is also possible to fuse the estimated depth from focus (DFD) depth map with a stereo map, e.g., via weighted fusion.[13] Ref. 14 uses two stereo pairs, each having a different focus, to separately model the depth map and the defocused image as Markov random fields. Finally, it is also possible to conduct a temporal defocus analysis method to recover a depth map by estimating the defocus kernel of the projector.[15] Our DFSI technique differs from these DFD techniques in that: 1. we only use two images, and 2. we uniformly model and solve for defocus blurs and scene depth.

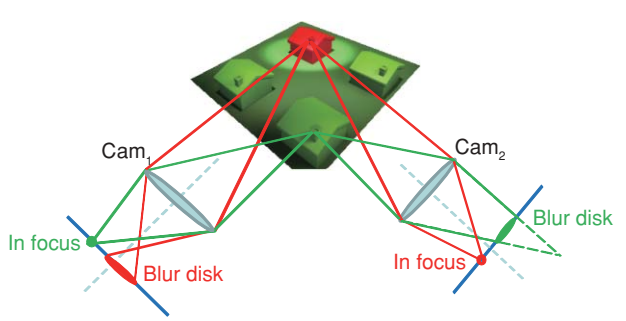


Fig. 2 Each image in a DFSI pair focuses at different scene depths. A 3-D point focused in one image will appear defocused in the other.

2.2 *Specially Designed Apertures*

In the emerging field of camera imaging, several researchers have suggested that replacing regular circular-shaped apertures with specially designed ones has many benefits. For example, coded apertures[4,5,16] change the frequency characteristics of defocus blur and use special deconvolution algorithms to reduce blurs and estimate scene depth from a single shot. A color-filtered aperture[2] divides the aperture into three regions, through which only light in one of the RGB color bands can pass. Out-of-focus regions will lead to depth-dependent color misalignments, which can then be used to determine scene depth. For more details, we refer the readers to the survey of computational photography.[1] A downside of these approaches is the requirement of modifying the camera optical system, whereas our DFSI technique does not require any modification as such.

2.3 *Defocus Matting*

One of the target applications of DFSI is automatic image matting. Image matting decomposes an observed image I into a foreground object image F, a background image B, and an alpha matte α. We refer the readers to Ref. 17 for an extensive survey on the state of the art approaches. While most matting schemes require user inputs in the form of scribbles or trimaps, several methods based on specially designed apertures[2] or apparati[9] have been recently proposed for automatically extracting matte. Defocus video matting[9] uses synchronized cameras to capture multiple images with different focuses from the same viewpoint. Light field matting synthesizes a defocused view using an array of cameras.[18] Our DFSI technique is able to automatically extract mattes from a pair of images captured by a hand-held camera.

2.4 *Multifocus Fusion*

Finally, our work is also related to methods for generating a multifocus photomontage. Ref. 4 uses coded apertures coupled with special deconvolution algorithms based on sparse-prior to recover an all-in-focus image for refocusing. Since they use regular circular-shaped apertures, their technique cannot fully recover the high-frequency components in the defocused images. Alternatively, digital photomontage systems[19] provide users with an interactive interface to highlight the desired regions in different images and then automatically fuse the selected regions. Dai and Wu used image matting techniques to iteratively recover the foreground and background layer from a partial blur image.[20] It is also possible to capture a sequence of images from the same viewpoint with varying focus and then merge them to synthesize an extended depth of field.[21] These systems either

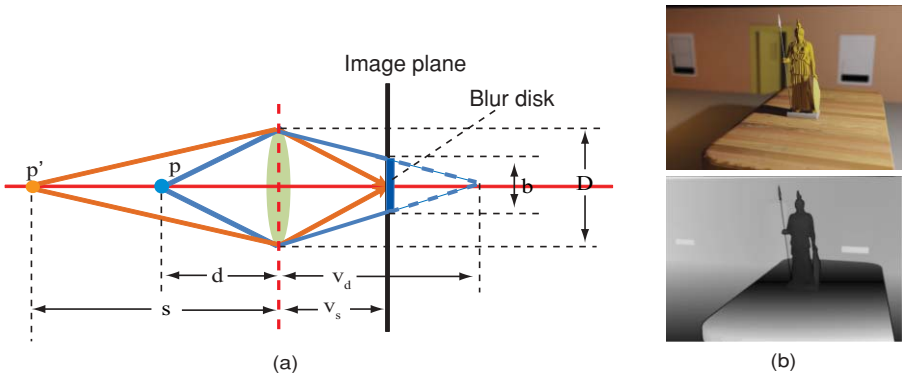


Fig. 3 Defocus kernel map (DKM). (a) The blur disk diameter *b* depends on the aperture size and the depth of the scene point. (b) Shows a sample DKM of a synthetic scene.

rely on user inputs and graph-cut optimization to segment the desired regions, or require viewpoint registrations, whereas DFSI can automatically segment and fuse the in-focus regions with smooth transitions.

3 Defocus Kernel Map

Similar to Refs. 11 and 14 and 22–26, we use Gaussian point spread functions (PSFs) to effectively approximate defocus blurs. Recent papers on coded apertures[4] have shown that other types of PSFs may be more suitable for reducing blur kernels. We, however, choose to use the Gaussian PSF for its simplicity in modeling the DKM. In fact, we derive the DKM constraint to directly correlate scene depth with Gaussian kernel sizes. In this work, the defocus blur at every pixel p is modeled as:

$$I(p) = I_0(p) \otimes b[d(p), c], \tag{1}$$

where $\otimes$ is the convolution operation. $I(p)$ is p's intensity after the defocus blur. $I_0(p)$ is p's intensity in the all-focus image. $b[d(p), c]$ is the blur kernel at p and is a function of p's depth $d(p)$ and the camera parameters c (i.e., camera aperture size and focal length). In this work, we call b the defocus kernel map (DKM).

In DFSI, since we only vary the scene focus while fixing the aperture size and the focal length, c remains constant and hence we can use $b(p)$ to represent the blur disk size at every pixel p. We first derive $b(p)$ in terms of c and $d(p)$. Assume the camera uses a thin lens of focal length f, aperture size D, and f-number $N = f/D$, and its image plane is positioned at v_s away from the lens to focus objects at depth s from the lens. s and v_s then satisfy the thin lens equation:

$$\frac{1}{s} + \frac{1}{v_s} = \frac{1}{f}. \tag{2}$$

Consider an arbitrary scene point p lying at depth $d(p)$; its image will focus at $v(p) = f \cdot d(p)/[d(p) - f]$ by the thin lens equation. If $d(p) < s$, then p's image will lie behind the image plane shown in Fig. 3(a). Thus, p will cast a blur disk of diameter $b(p)$, where

$$b(p) = \frac{[v(p) - v_s] \cdot D}{v(p)}. \tag{3}$$

Substituting $v(p)$ and v_s with $d(p)$ and s using the thin lens equation, we have:

$$b(p) = \alpha \frac{|d(p) - s|}{dp \cdot (s - \beta)}, \tag{4}$$

where $\alpha = f^2/N$, and $\beta = f$. Eq. (4) computes the blur disk size at every pixel in terms of its scene depth, and we call it the *defocus constraint.*

3.1 *Disparity Defocus Constraint*

Next, we derive the *defocus constraint* in terms of the disparity map in DFSI. (The disparity map here refers to the one computed between the corresponding all-focus images.) DFSI uses a pair of images I_1 and I_2 captured with the same f-number and focal length, but focused at different scene depths. For every pixel p in I_1, its disparity $\gamma(p)$ (with respective to I_2) can be computed from its depth $d(p)$ as $\gamma(p) = K/d(p)$, where K is a function of the camera baseline and intrinsic parameters, and is constant for all pixels. Similarly, we can map the in-focus scene depth s in I_1 to its disparity γ_s. Substituting $d(p)$ and s with $\gamma(p)$ and γ_s, in the *defocus constraint* Eq. (4), we have:

$$b(p) = \tilde{\alpha} \frac{|\gamma_s - \gamma(p)|}{\tilde{\beta} - \gamma_s}, \tag{5}$$

where $\tilde{\alpha} = \alpha/\beta$ and $\tilde{\beta} = k/\beta$. We call Eq. (5) the *disparity defocus constraint.* Similarly, we can compute the DKM b of I_2 with respect to I_1. For the rest of the work, we refer by default, the DKM to the one associated with I_1.

4 Defocused Stereo Matching

In this section, we show how to simultaneously recover the DKMs and the disparity map. We assume that each image in a DFSI pair focuses at some scene objects/features. This is a common practice in photography, especially when one uses a hand-held camera. Our algorithm starts with finding SIFT feature correspondences[27] between the DFSI pair, and apply Ref. 28 to rectify the images. Next, we extract the salient features in each rectified image and estimate their initial disparity values to recover the camera parameters $\tilde{\alpha}$ and $\tilde{\beta}$ (see Sec. 4.1). We then integrate the defocus kernel map estimation with the disparity map solution process using the *pair-wise defocus constraint.* Finally, we iteratively refine the camera parameters, the DKMs, and the disparity map. Fig. 4 illustrates the processing pipeline of our algorithm.

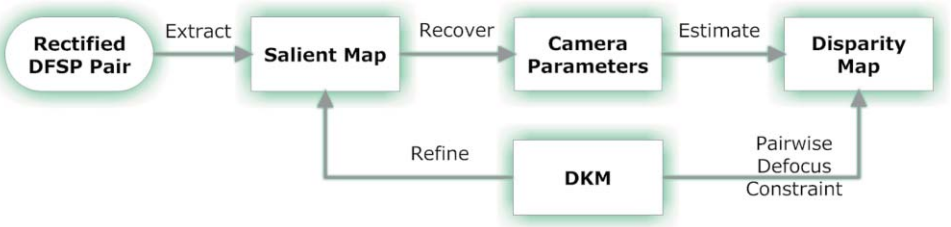


Fig. 4 The processing pipeline of the DFSI technique.

4.1 *Recovering Camera Parameters*

We first develop a simple but effective algorithm to recover the camera parameters. Since we have two unknowns $\tilde{\alpha}$ and $\tilde{\beta}$, we use two disparity defocus constraints to solve for them. To do so, we estimate the disparity defocus constraints for the in-focus pixels Ω_1 and Ω_2 in I_1 and I_2, respectively. Assuming Ω_1 has disparity γ_1 and casts a blur disk of size b_2 in I_2, and Ω_2 has disparity γ_2 and casts a blur disk of size b_1 in I_1, the two *disparity defocus constraints* are:

$$\begin{cases} b_1 = \tilde{\alpha}\dfrac{|\gamma_1 - \gamma_2|}{\tilde{\beta} - \gamma_1} \\ b_2 = \tilde{\alpha}\dfrac{|\gamma_2 - \gamma_1|}{\tilde{\beta} - \gamma_2} \end{cases}. \tag{6}$$

Since a DFSI pair focus at different scene depths, $\gamma_1 \neq -\gamma_2$ and Eq. (6) are nondegenerate. Our goal is to find γ_1, γ_2, b_1, and b_2, and solve for $\tilde{\alpha}$ and $\tilde{\beta}$.

To find γ_1 and γ_2, we first compute the salient features by applying a high-pass filter on I_1 and I_2. To minimize outliers, we blur each image I_i using a small Gaussian kernel and then subtract the blurred image from I_i. We use Gaussian kernels, as they are coherent with the defocus blur model and effectively suppress aliasing artifacts such as ringing. We then threshold the high-pass filtered images to obtain initial salient maps, as shown in Figs. 5(a) and 5(b). We also use the graph-cut algorithm to compute the initial disparity map, and assign the computed disparity value to all salient feature points. We assume all salient features in their corresponding images correspond to the same depth, and hence have the same disparity. To remove outliers, we use the median disparity value of feature points in I_1 and I_2 as γ_1 and γ_2.

To find b_1 and b_2, we locate the corresponding pixels and exhaust all possible kernel sizes b. To find b_2, for each feature point p in the salient map of I_1, we apply Gaussian blur of size b_2 at pixel p and compute the difference between p and $q = p + \gamma_1$ in I_2. We find the optimal b_2 that minimize the summed squared difference for all salient features in I_1. Similarly, we obtain the optimal b_1. Finally, we combine b_1 and b_2 with γ_1 and γ_2 to solve for $\tilde{\alpha}$ and $\tilde{\beta}$ using disparity defocus constrain Eq. (5).

4.2 *Defocus Kernel Map Disparity Markov Network*

Classical stereo matching methods model the disparity map as a Markov random field (MRF). The problem can be treated as assigning a label $\gamma: P \rightarrow \Gamma$, where P is the set of all pixels and Γ is a discrete set of labels corresponding to different disparities. Graph cut[29–31] and belief propagation[32–34] can be used to find the optimal labeling. Since the DKMs can be directly derived from the disparity map [Eq. (5)], they share similar MRF properties as the disparity map, i.e., they are smooth except for crossing a boundary. Therefore, we can integrate DKM estimation with the graph-cut-based disparity map estimation process.

We define the energy function E as:

$$E(\gamma) = \sum_{p \in I_1} E_r[p, \gamma(p)] + \sum_{p_1, p_2 \in N} E_s[\gamma(p_1), \gamma(p_2)], \tag{7}$$

where the data penalty term E_r describes how well the disparity γ fits the observation, the smoothness term E_s encodes the smoothness prior of Γ, and N represents the pixel neighborhood in image I_1.

In our implementation, we use the similar smoothness term as in Ref. 31. The data penalty $E_r[p, \gamma(p)]$ measures the appearance consistency between pixel p in I_1 and pixel $q = p + \gamma(p)$ in I_2. Recall that I_1 and I_2 have different foci. Thus, even with the correct disparity γ, the appearance of p and q may appear significantly different due to defocusing. Therefore, we cannot directly compare the intensity between $I_1(p)$ and $I_2(q)$.

Note that given γ and the recovered camera parameters $\tilde{\alpha}$ and $\tilde{\beta}$, we can directly compute the defocus blur kernel b_p and b_q at pixel p and q using Eq. (5). Assuming p is less blurry than q, we can apply additional Gaussian blur G_σ to p in I_1, and then compared the blurred result with q. We call the resulting images I_1^* and I_2^* an *equally defocused* pair:

$$\begin{aligned} I_1^*(p) &= \begin{cases} I_1(p) \otimes G_\sigma, & b_p < b_q \\ I_1(p), & \text{otherwise} \end{cases}. \\ I_2^*(q) &= \begin{cases} I_2(q) \otimes G_\sigma, & b_q < b_p \\ I_2(q), & \text{otherwise} \end{cases}. \\ \sigma &= \sqrt{|b_p^2 - b_q^2|}. \end{aligned} \tag{8}$$

Finally, we define E_r as

$$E_r[p, \gamma(p)] = \min\{0, [I_1^*(p) - I_2^*[p + \gamma(p)]]^2 - K_2\}, \tag{9}$$

where the truncation threshold K_2 is used to reduce noise and remove outliers. We use graph cut to solve for the optimal disparity map and apply the disparity defocus constraint for computing the DKMs. In theory, we could further improve the disparity map by modeling the occlusion boundaries.[31,33] In practice, we find it sufficient and robust to use the estimated DKM for segmentation in the following sections.

Once we obtain the disparity map and DKMs, we can refine the salient feature maps by removing points whose disparity deviates from the median disparity of all feature points. We then re-estimate the camera parameters (see

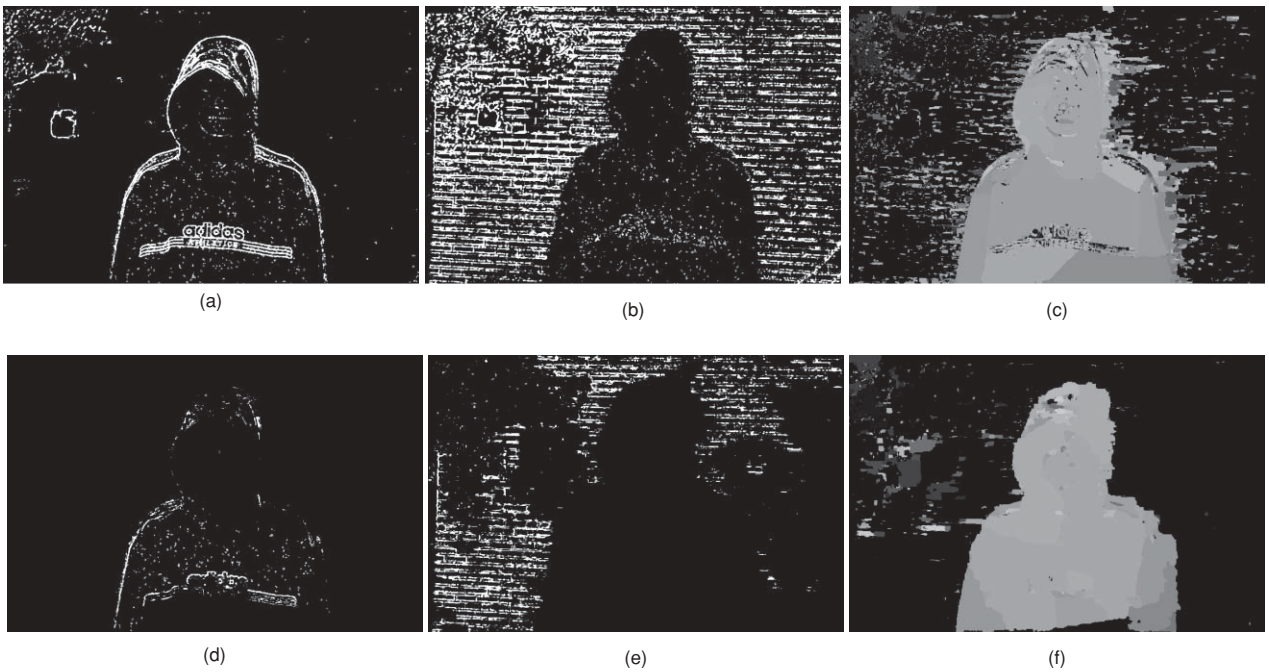


Fig. 5 The recovered disparity map. (a) and (b) show the initial salient feature maps of the DFSI pair in Fig. 7. (c) shows the initial disparity map. (d) and (e) show the refined salient feature map after pruning the outliers. (f) shows the recovered disparity map after two iterations.

Sec. 4.1). We repeat this iterative refinement two to three times. Fig. 5(f) shows the final estimated disparity map.

4.3 *Defocus Kernel Map Based Segmentation*

The recovered DKMs can be used to robustly segment the in-focus region in each DFSI image. Specifically, we treat the segmentation problem as a labeling problem on the DKM and use only two labels, S for the foreground and T for the background. To find the optimal labeling, we define the energy function E as:

$$E(L) = \lambda \cdot \sum_{p \in P} E_r[p, L(p)] + \sum_{p,q \in N} E_b[p, q, L(p), L(q)], \tag{10}$$

where P represent all pixels in the image, N represents the pixel neighborhood, and $L(p)$ is the labeling at p. The non-negative coefficient λ specifies a relative importance of region penalty E_r and boundary penalty E_b. To model discontinuities in labeling, we model E_b as,

$$E_b[p, q, L(p), L(q)] = \begin{cases} \dfrac{\exp\{-[I(p) - I(q)]^2\}}{\|p - q\|}, & L(p) \neq L(q). \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Unlike previous segmentation methods[35,36] that define the region penalty E_r using the histogram distribution or Gaussian mixture models from user-specified foreground/ground samples, we directly compute E_r in terms of the defocus kernel map B. Notice that a smaller defocus kernel b corresponds to a higher likelihood that the pixel is in-focus. Thus, we simplify E_r as

$$E_r(p, S) = \begin{cases} K_3 \cdot (M - b_p), & b_p < M \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

$$E_r(p, T) = \begin{cases} b_p - M, & b_p \geq M \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

where M is the maximum size of the circle-of-confusion that would be considered in focus. K_3 is a positive scaling factor for balancing the region penalty between the foreground and background. In our experiments, we set $M = 5$ and $K_3 = (b_{\max} - M)/M$, where $b_{\max}$ corresponds to the maximum size of the blur disk. A sample segmentation result is shown in Sec. 5.1.

5 Applications

In this section, we demonstrate how to apply DFSI for various imaging applications, including low-light imaging, automatic defocus matting, and multifocus photomontage.

5.1 *Low-Light Imaging*

Capturing high quality images under low light is a challenging problem. Images captured with a regular aperture and shutter setting are commonly underexposed and noisy, as shown in Fig. 6(a). One possible solution is to use slow shutters. However, slow shutters can cause significant image blur due to scene and/or camera motions. In DFSI, slow shutters can be particularly problematic, as we use a hand-held camera to capture the images, as shown in Fig. 6(c). The resulting motion blur is difficult to correct, as it is not spatially invariant.[37,38] Another possible solution is to denoise the images, e.g., via principal component/tensor analysis.[39] However, a large number of images (~20) are often required

(a) (b) (c)

Fig. 6 Problems with low-light imaging. (a) An image captured with a small aperture and fast shutter appears underexposed. (b) shows the histogram stretched result of (a). (c) An image captured using a small aperture and slow shutter exhibits severe motion blur.

for robust denoising. Alternatively, we can use wide apertures in place of slow shutters, which would lead to shallow DOFs and only part of the scene can be clearly focused, as shown in Figs. 7(a) and 7(b).

In this section, we use DFSI pairs to enhance low-light imaging. We place two synchronized digital SLR cameras adjacent to each other. The two cameras use an identical aperture size, with one focusing on the foreground and the other on the background. Instead of deblurring the images, we explore effective fusion methods to combine in-focus regions in both DFSI images.

Recall that we use the recovered DKMs to segment the in-focus region (see Sec. 4.3). A naive approach is to directly warp the segmented region using the disparity map. In practice, the estimated disparity map can still be noisy and discontinuous along the segmentation boundaries. Hence, directly warping the boundary will cause jigjagged or scattered edges in the final fused image. To resolve this issue, we develop a *Snake*-based[40] contour warping algorithm.

Given the boundary V_1 of the in-focus region in I_1, our goal is to find its optimal *target* boundary V_2 in I_2, so that when V_1 is warped to V_2 using the recovered disparity map, it will be continuous with respect to the background in I_2. To do so, we define the energy function for all pixels $\tilde{p}$ on contour V_2 as:

$$E_{\text{contour}} = \sum_{\text{all}\,\tilde{p},\,\tilde{p}\in V_2} [E_{\text{image}}(\tilde{p}) + \zeta \,.\, E_{\text{shape}}(\tilde{p})], \tag{14}$$

where E_{image} describes appearance similarity and E_{shape} describes the shape similarity.

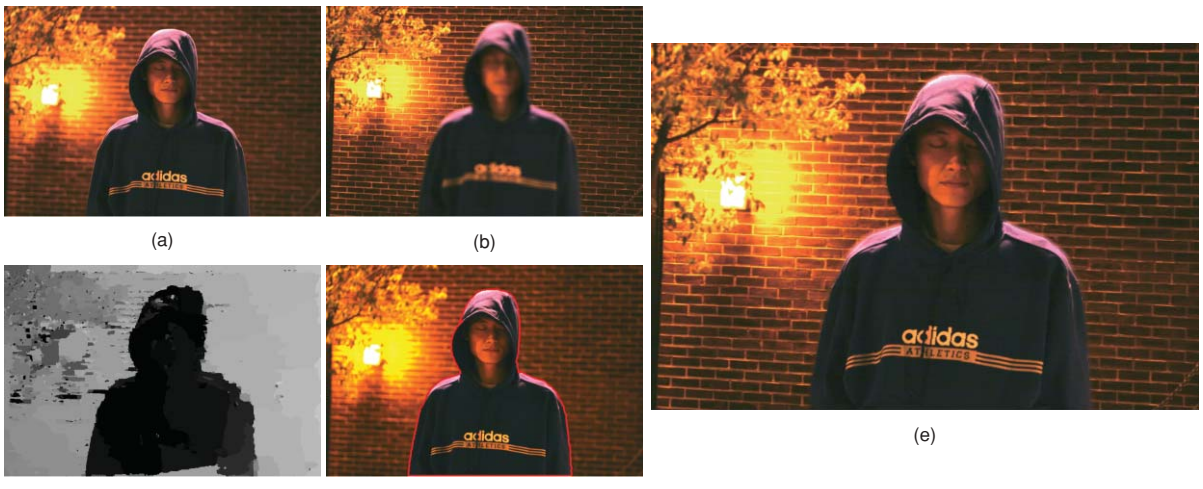
Recall that we cannot directly compare the intensity between pixel p on V_1 and p on V_2, since they have different defocus levels. Therefore, we first use the recovered DKMs to compute the equally defocused image pair I_1^* and I_2^* using Eq. (8). We then measure the similarity between p and $\tilde{p}$ using I_1^* and I_2^* as:

$$E_{\text{image}}(\tilde{p}) = |F[I_1^*(p)] - F[I_2^*(\tilde{p})]|, \tag{15}$$

where $F[I_1^*(p)] = (I_1^* \otimes G)(p), F[I_1^*(\tilde{p})] = (I_2^* \otimes G)(\tilde{p})$, and G is a Gaussian kernel that serves as a weighting function.

In addition, we enforce the shape similarity between the two contours. Specifically, we measure the similarity of the first and second order differential geometry attributes between the corresponding point p and $\tilde{p}$ on V_1 and V_2 as:

$$E_{\text{shape}}(\tilde{p}) = \|V_1'(p) - V_2'(\tilde{p})\| + \|V_1''(p) - V_2''(\tilde{p})\|, \tag{16}$$



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

Fig. 7 DFSI for low-light imaging. (a) and (b) show a DFSI pair. We use (c) the DKM to segment the in-focus regions (d) in the first image. Finally, we warp the segmented in-focus regions to the second image using the estimated disparity map to form a nearly all-in-focus image in (e).

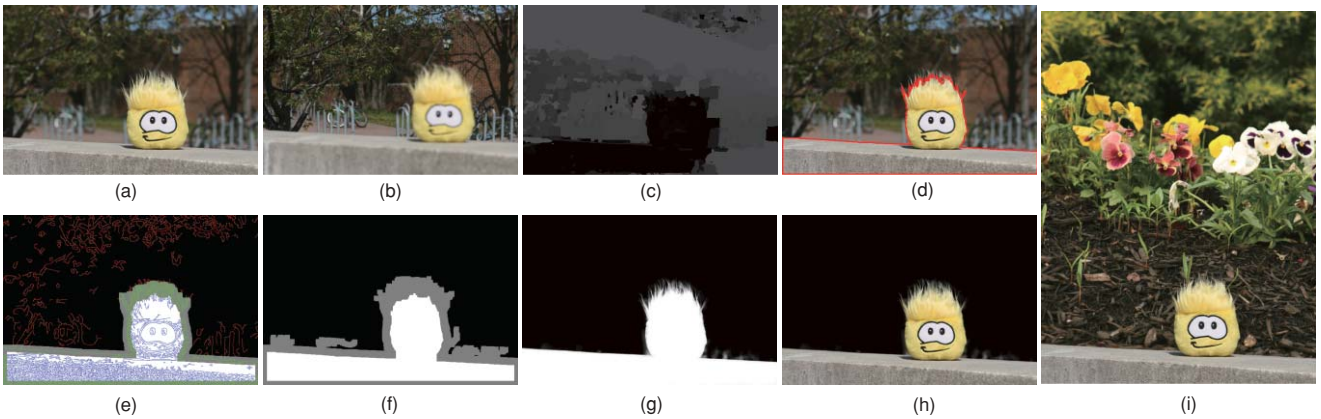


Fig. 8 DFSI for alpha matting. (a) and (b) show a DFSI pair of a children's toy scene, with the left image focusing on the penguin and the right on the background tree. Notice the strong parallax between the images. We recover the DKM in (c) and use it to segment the foreground in (d). (f) We create a trimap using (e) morphologic operations based on edge heuristics. (g) and (h) show the recovered alpha matte and foreground using robust matting.[46] (h) We create a composite image using a new background.

and we apply a Greedy algorithm similar to *Snakes*[40] to find the optimal boundary in V_2.

The computed disparity inside the segmented region can also be noisy. Therefore, we only use the disparity estimation on the boundary pixels and smoothly interpolate the disparity value for the interior pixels. Finally, we use the interpolated disparity map for warping all pixels inside the in-focus region to the final fused image. Fig. 7(e) shows such an example.

5.2 *Automatic Defocus Matting*

Next, we apply the DFSI technique for automatically extracting the alpha matte. The problem of image matting has been studied for decades. Recently, several multiimage-based matting methods have been proposed to automatically extract the matte. For example, defocus video matting[9] uses a special imaging system to capture multiple images of the scene from the same viewpoint and with different foci. It then automatically classifies the image into a foreground Ω_F, a background Ω_B, and an unknown region Ω_U by analyzing the defocus blur. However, multiimage-based methods are sensitive to calibration errors, camera shake, and foreground motions.

We use the recovered DKMs from the DFSI pair to automatically generate the trimap. For trimap-based matting schemes, the quality of the alpha matte relies heavily on the estimation accuracy of the unknown region Ω_U. One way to generate Ω_U is to erode the estimated foreground and background as:

$$\Omega_U = \overline{\text{erode}(\Omega_F, \tilde{b}) \cup \text{erode}(\Omega_B, \tilde{b})}, \tag{17}$$

where erode is a morphological operator[41] and $\tilde{b}$ is the width of erosion. For DFSI, we can compute Ω_F and Ω_B by the DKM-based segmentation and $\tilde{b}$ by averaging the blur kernel of all pixels in Ω_B.

Such a trimap generation method works well when the segmented foreground boundary is relatively accurate. However, for DFSI, the boundary of the in-focus region is extracted based on the estimated disparity map. In the presence of fuzzy boundaries, it is difficult to distinguish defocus blur from fuzziness, and the resulting foreground boundary can be inaccurate, as shown in Fig. 8(d).

We present a new scheme for generating the trimap by exploiting edge continuities. The key observation here is that, for fuzzy objects, strong edges along the foreground boundary should be considered part of the unknown region. Therefore, we grow the unknown region along the edge directions. Specifically, we first use the Canny edge detector[42] to locate the strong edges, then dilate these edges and combine them with Ω_U to form the extended unknown region Ω_U^* as

$$\begin{aligned} \Omega_1 &= \text{dilate}(\Omega_U, 2\tilde{b}) \cap \text{dilate}[\text{Canny}(I), \tilde{b}], \\ \Omega_U^* &= \Omega_1 \cup \Omega_U, \\ \Omega_U^* &= \Omega_F \cap \overline{\Omega_U^*}. \end{aligned} \tag{18}$$

Once we obtain the trimap, we use robust matting to estimate the alpha matte, the foreground, and the background. Figs. 8 and 9 show the automatic defocus matting results using our approach.

5.3 *Multifocus Photomontage*

Finally, we apply DFSI for creating a multifocus photomontage that virtually focuses at multiple scene depths. Synthesizing a multifocus photomontage can benefit many applications. For example, in confocal microscopy, only tissues lying near the scanning layer can be clearly imaged, and it is highly desirable to combine the in-focus regions from all layers. Another interesting application is to synthesize novel defocusing effects, e.g., by focusing on both the foreground and background while defocusing on the in-between ground.

Our DFSI-based multifocus photomontage differs from existing multifocus fusion approaches[12,19] in several ways. First, we use images captured from different viewpoints, whereas existing approaches assume that the images are captured from or can be warped to the same viewpoint. Second, most photomontage techniques require using a large number of images to accurately identify the in-focus regions while we only use a pair of images. Finally, unlike digital photomontage[19] that relies on user inputs, our DFSI-based photomontage method is fully automatic.

Given a DFSI pair, we start by segmenting the in-focus region using the recovered DKMs. In low-light imaging (see Sec. 5.1), we directly warp the segmented in-focus region

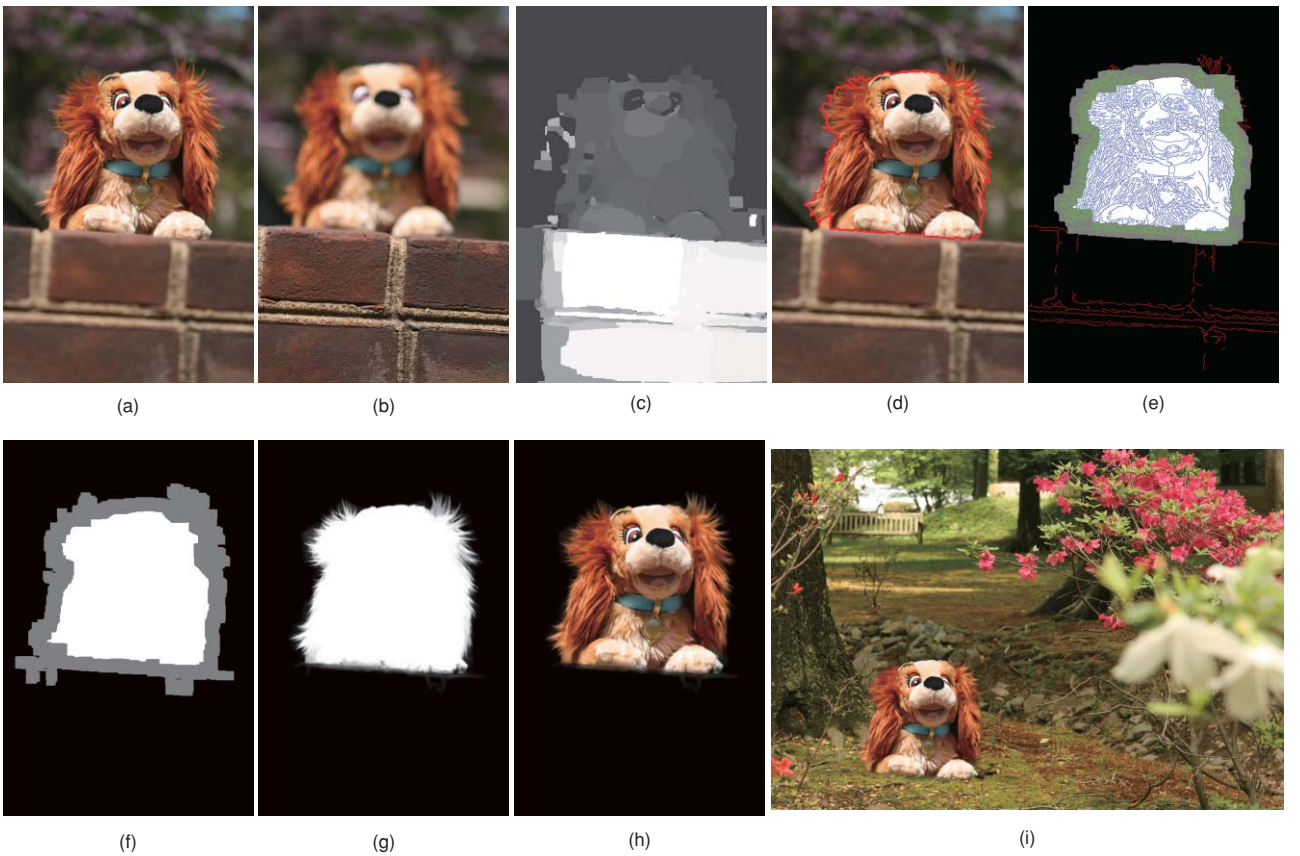


Fig. 9 DFSI for background matting. (a) and (b) show a DFSI pair of a toy dog scene, with the first image focusing at the foreground bricks and the second at the middle-ground toy dog. (c) through (f) show the recovered disparity map, segmented boundary, edge heuristic overlaid with a trimap by Eq. (17), and our trimap, respectively. (g) and (h) show the extracted alpha matte and foreground. (i) shows the composite result.

from one image to the other. We assume that scene objects lie on different depth layers and rely on Snake for locating the occlusion boundary. For multifocus photomontage, scene depth can vary smoothly (e.g., the mulch ground in Fig. 10). Therefore, directly warping the foreground region will lead to sudden changes of blurriness across the warping boundary, as shown in Fig. 10(g).

To resolve this issue, we compute the DKM for each image in the DFSI pair to determine the defocus blur kernel size near the warped boundaries. We then blur the warped in-focus boundary region accordingly to maintain smooth transitions. Specifically, given the warped boundary, we first erode and dilate the boundary Ω by a fixed width to form a band. Our goal is to determine how much blur should be applied to each pixel inside the band after warping. We call this process adaptive boundary blurring.

To maintain the same sharpness/blurriness outside the band, we set the adaptive blur kernel size to be zero on both the interior and exterior boundaries Ω^+ and Ω^- that are obtained by erosion and dilation, similar to the way we obtain the trimap for defocus matting (see Sec. 5.2). We determine the adaptive blur kernels on Ω from the estimated DKM and use natural neighbor coordinates[43] to interpolate the missing pixels inside the band. Figs. 10(e) and 10(j) show the warped boundary and the adaptive blur kernel map. Figs. 10(f) and 10(h) compare the results between direct warping and adaptive blurring. Using adaptive blurring, our photomontage technique is able to maintain smooth transitions from the in-focus region to the out-of-focus region.

6 Results

In our experiments, all images were captured using a Canon Rebel XTi digital SLR camera with a Canon EF 24–70-mm f/2.8L lens. We first use DFSI to recover the DKMs and the disparity map. We capture a DFSI pair with Av 2.8 and Tv 1/8, as shown in Figs. 7(a) and 7(b). We start with computing the initial salient feature maps [Figs. 5(a) and 5(b)] and the disparity map [Fig. 5(c)] for estimating the camera parameters. Figs. 5(d), 5(e), and 5(f) show the refined salient feature maps and the disparity map after two iterations. Our iterative optimization effectively removes outliers in salient feature maps and significantly improves the disparity map.

In Fig. 6, we demonstrate using the DFSI technique for acquiring images under low light. An image captured under a regular aperture and shutter setting (Av 11, Tv 1/8) appears underexposed, as shown in Fig. 6(a). Gamma correction and histogram equalization can be used to enhance contrast. However, the resulting image is still noisy due to insufficient lighting [Fig. 6(b)]. To increase exposures, we use a slower shutter (Tv 2) [Fig. 6(c)]. However, the image contains severe motion blurs due to handshakes. Next, we capture a

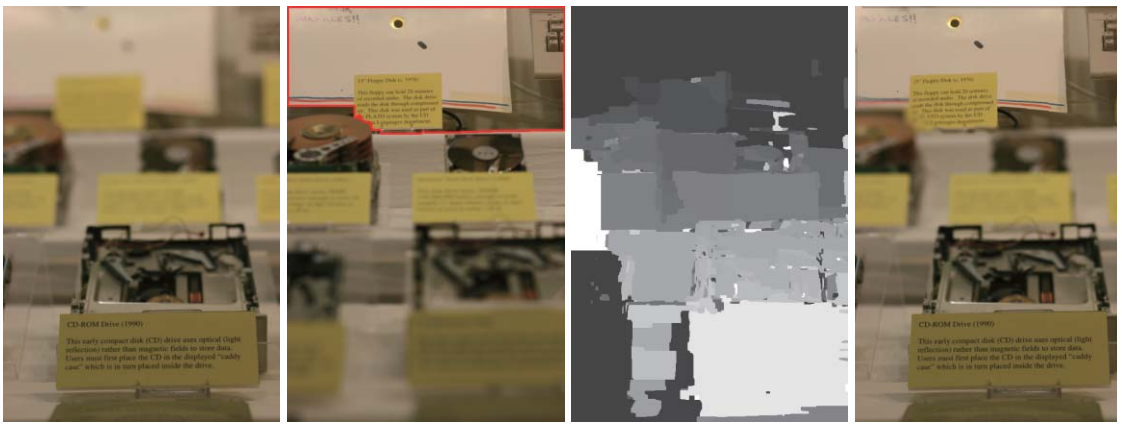
Fig. 10 A multifocus photomontage on a continuous scene. (a) and (b) show a DFSI pair of a garden scene. The first image focuses on the foreground sprouts and the second on the middle-ground flower. (c) and (d) are the recovered disparity map and the segmentation result. (e) shows the warped boundary using the Snake-based algorithm. (f) shows the fusion result by directly warping the segmented foreground to the second image. Notice that the blurriness changes abruptly across the boundary in (g). (h) maintains a smooth boundary by blurring the segmentation boundary using (j) the adaptive kernel map and (i) the DKM.

pair of DFSI images [Figs. 7(a) and 7(b)]. We recover the DKM [Fig. 7(c)] and use it to segment the in-focus region. Finally, we warp the segmentation result to Fig. 7(e). The final synthesized image preserves final details with minimal noise and motion blur.

In Fig. 8, we demonstrate automatic defocus matting of a children's toy scene using DFSI. Figs. 8(a) and 8(b) were captured with a wide aperture and fast shutter (Av 6.3 and Tv 1/640). Fig. 8(a) focuses at the foreground of the toy and Fig. 8(b) at the background trees. To extract the matte, we recover the DKM [Fig. 8(c)] and use it to segment the foreground [Fig. 8(d)]. Since the right half of the stone bench appears textureless due to slight defocus blur, its DKM is not accurate. However, the DKM is only used to initialize the segmentation process; therefore we are still able to accurately segment the foreground region. Fig. 8(e) shows the edge heuristics for trimap computing: all edges are detected by the Canny operator and overlaid with the trimap estimated using Eq. (17), where red, green, and blue edges lie in the estimated background, unknown, and foreground regions, respectively. Fig. 8(f) shows the final trimap created by the boundary growing algorithm (see Sec. 5.2). Figs. 8(g) and 8(h) show the recovered alpha matte and foreground using robust matting. In Fig. 8(i), we create a composite image by using a new background. In Fig. 9, we show another example using DFSI defocus matting in a toy dog scene.

In Fig. 10, we apply DFSI for a multifocus photomontage. We capture a DFSI pair of a garden scene using F2.8. The first image focuses on the foreground sprouts and the second at the middle ground flower. Fig.10(c) shows the recovered disparity map. We use the DKM to segment the foreground region, as shown in red in Fig. 10(d). Next, we warp the foreground to the second image [Fig. 10(f)]. However, directly warping the foreground region incurs abrupt changes of blurriness across the warping boundary. Therefore, we further compute an adaptive kernel map [Fig. 10(j)] from the foreground boundary. We use the second DKM [Fig. 10(i)] to adaptively blur the boundary. Fig. 10(g) shows the close-up views of the fusion results with and without adaptive blur.

In Figs. 11(a) and 11(b), we create a multifocus photomontage using DFSI on a computer museum scene where the images focus at different tags. Fig. 11(b) shows the background segmentation result for the second view. We then warp the background to the first view and apply adaptive blur to maintain smooth boundary transition. Fig. 10(d) illustrates a novel multifocus photomontage effect by keeping



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

Fig. 11 A multifocus photomontage on a computer museum scene. The first image (a) in the DFSI pair focuses on the foreground tags and the second (b) at the background tags. (c) shows the estimated disparity map. The segmented background boundary is shown as red in (b). (d) shows the final fused image with a novel multifocus effect, where only the middle ground is defocused.

both the foreground and background in focus and the middle ground out of focus.

7 Discussions and Future Work

We demonstrate a novel dual-focus stereo imaging technique. DFSI captures a pair of images from different viewpoints and at different foci, but with identical wide aperture size. Table 1 summaries the major difference between DFSI and other related work. Compared with coded aperture imaging,[4] DFSI does not require using specially designed optical systems or modifying the camera. Furthermore, DFSI can be directly applied to enhance low-light imaging by coherently increasing the aperture size in both images, which is hard to achieve using coded apertures. Compared with color-filtered aperture imaging,[2] DFSI avoids complex color calibration procedures and works robustly in the presence of single-colored scene objects.

Table 1 Comparisons with state of the art methods.

	Number of Images	Coaxial	Special equipment	Output
Ref. 9	3	Yes	Optical bench and beam-splitters	Matte
Ref. 18	8	No	1×8 camera array	Matte
Ref. 4	1	No	Coded aperture masks	Depth map
Ref. 2	1	No	Color filtered aperture	Depth map and matte
Ref. 12	451 to 793	Yes	None	Depth map
DFSI	2	No	None	Depth map and matte

For extracting the alpha matte, DFSI shares similarities with the defocus video matting system. However, DFSI does not require all images to be captured from the same viewpoint, and hence avoids using a special apparatus. While parallax between images can cause problems in Ref. 9, DFSI effectively uses parallax for recovering the DKMs. Finally, compared with multicamera video matting, DFSI only uses a pair of images, whereas the system in Ref. 18 requires using a large number of images to synthesize the defocusing effect.

DFSI, however, has several limitations. In our implementation, we use spatially variant Gaussian PSFs to approximate defocus blur kernels. Since the kernels are no longer Gaussian near occlusion boundaries, we are unable to accurately estimate scene depth for the boundary pixels, and our fusion technique generates visual artifacts such as bleeding and discontinuity. These artifacts have also been observed in the gathering methods[23,44] used to render DOF effects in computer graphics. The scattering method,[45] in theory, can help reduce the bleeding/discontinuity artifacts. However, it requires highly accurate scene geometry to avoid aliasing. Since our DFSI technique only uses two images, our estimated scene depth map cannot reach the accuracy level.

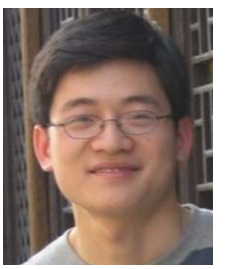
Another major limitation of our approach is that we rely on the disparity map solution for recovering the DKMs. In our approach, we compute the disparity labeling by blurring the relatively sharp pixel with the optimal kernel to match the blurrier one. Levin et al.[4] have shown that due to the frequency characteristics of the circular aperture, multiple kernels exist that produce similar blurry results. Therefore, our disparity map may introduce large errors in regions that are defocused in both images. For future applications such as 3-D reconstruction, we plan to explore possible combinations of image statistics and DFSI to more accurately recover scene depth. Finally, all examples shown in this work paper are captured by still cameras. It would be a natural extension to our framework to use a pair of DFSI video cameras for acquiring low-light videos and performing video matting.

Acknowledgments

Li and Yu were partially supported by the National Science Foundation under grants MSPA-MCS-0625931 and IIS-CAREER-0845268.

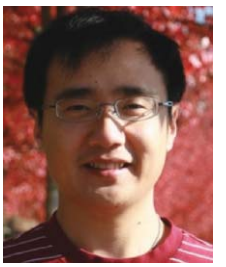
References

1. R. Raskar, J. Tumblin, A. Mohan, A. Agrawal, and Y. Li, "Computational photography," in *Proc. Eurographics STAR* (2006).
2. Y. Bando, B. Y. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," *ACM Trans. Graph.* **27**(5), 1–9 (2008).
3. P. Green, W. Sun, W. Matusik, and F. Durand, "Multi-aperture photography," *SIGGRAPH '07*, p. 68 (2007).
4. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a aperture," in *SIGGRAPH '07*, p. 70 (2007).
5. A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," in *SIGGRAPH '07*, p. 69 (2007).
6. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Tech. Rep. CSTR 2005-02, Stanford University Computer Science* (Apr. 2005).
7. R. Raskar, K. H. Tan, R. Feris, J. Yu, and M. Turk, "Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging," *ACM Trans. Graph.* **23**(3), 679–688 (2004).
8. R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," *ACM Trans. Graph.* **25**(3), 795–804 (2006).
9. M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand, "Defocus video matting," in *SIGGRAPH '05*, pp. 567–576 (2005).
10. C. Frese and I. Gheta, "Robust depth estimation by fusion of stereo and focus series acquired with a camera array," *Multisensor Fusion Integration Intel. Syst.*, pp. 243–248 (2006).
11. J. Kim and T. Sikora, "Confocal disparity estimation and recovery of pinhole image for real-aperture stereo camera systems," *ICIP*, pp. 229–232 (2007).
12. S. W. Hasinoff and K. N. Kutulakos, "Confocal stereo," *Intl. J. Comput. Vision* **81**(1), 82–104 (2009).
13. N. Ahuja and A. L. Abbott, "Active stereo: integrating disparity, vergence, focus, aperture and calibration for surface estimation," *Patt. Anal. Mach. Intell.*, pp. 1007–1029 (Oct. 1993).
14. A. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, "Depth estimation and image restoration using defocused stereo pairs," *Patt. Anal. Mach. Intell.* **26**(11), 1521–1525 (Nov. 2004).
15. L. Zhang and S. Nayar, "Projection defocus analysis for scene capture and image display," *ACM Trans. Graph.* **25**(3), 907–915 (2006).
16. C. Zhou and S. Nayar, "What are good apertures for defocus deblurring?" *IEEE Intl. Conf. Comput. Photo.*, 1–8 (Apr. 2009).
17. J. Wang and M. Cohen, "Image and video matting: a survey," *Found. Trends Computer Graph. Vision* **3**(2), pp. 97–175 (2007).
18. N. Joshi, W. Matusik, and S. Avidan, "Natural video matting using camera arrays," *ACM Trans. Graph.* **25**(3), 779–786 (2006).
19. A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Trans. Graph.* **23**(3), 294–302 (2004).
20. S. Dai and Y. Wu, "Removing partial blur in a single image," *Computer Vision Patt. Recog.*, pp. 2544–2551 (2009).
21. S. W. Hasinoff and K. N. Kutulakos, "Light-efficient photography," in *ECCV '08: Proc. 10th Euro. Conf. Computer Vision*, 45–59 (2008).
22. H. Farid and E. P. Simoncelli, "Range estimation by optical differentiation," *J. Opt. Soc. Am.* **15**, 1777–1786 (1998).
23. S. Lee, G. J. Kim, and S. Choi, "Real-time depth-of-field rendering using anisotropically filtered mipmap interpolation," *IEEE Trans. Visual. Computer Graph.* **15**(3), 453–464 (2009).
24. A. P. Pentland, "A new sense for depth of field," *PAMI* **9**(4), 523–531 (1987).
25. M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *Int. J. Comput. Vision* **27**(3), 203–225 (1998).
26. Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," in *CVPR*, pp. 68–73 (1993).
27. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Comput. Vision* **60**(2), 91–110 (2004).
28. M. Pollefeys, R. Koch, and L. V. Gool, "A simple and efficient rectification method for general motion," *ICCV* **1**, 496–501 (1999).
29. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *PAMI* **23**(11), 1222–1239 (2001).
30. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *ICCV*, pp. 508–515 (2001).
31. V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *ECCV*, pp. 82–96 (2002).
32. P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," pp. 261–268 (2004).
33. J. Sun, Y. Li, S. B. Kang, and H. Y. Shum, "Symmetric stereo matching for occlusion handling," in *CVPR*, **2**, 399–406 (2005).
34. J. Sun, H. Y. Shum, and N. N. Zheng, "Stereo matching using belief propagation," in *PAMI* **25**(7), 787–800 (2003).
35. Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *IJCV* **70**(2),109–131 (2006).
36. C. Rother, V. Kolmogorov, and A. Blake, "'grabcut': interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.* **23**(3), 309–314 (2004).
37. R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.* **25**(3), 787–794 (2006).
38. Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.* **27**(3), 1–10 (2008).
39. L. Zhang, S. Vaddadi, H. Jin, and S. Nayar, "Multiple view image denoising," *Computer Vision Patt. Recog.*, pp. 1542–1549 (2009).
40. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV* **1**(4), 321–331 (1988).
41. R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Patt. Anal. Mach. Intell.* **9**(4), 532–550 (1987).
42. J. Canny, "A computational approach to edge detection," *IEEE Trans. Patt. Anal. Mach. Intell.* **8**(6), 679–698 (1986).
43. R. Sibson, "*A brief description of natural neighbour interpolation,*" in *Interpreting Multivariate Data,* V. Barnet, Ed., John Wiley and Sons, Chichester pp. 21–36 (1981).
44. J. Earl Hammon, "Practical post-process depth of field," *GPU Gems 3* **28**, 583–606 (2007).
45. M. Potmesil and I. Chakravarty, "A lens and aperture camera model for synthetic image generation," in *SIGGRAPH '81*, pp. 297–305 (1981).
46. J. Wang and M. Cohen, "Optimized color sampling for robust matting," in *CVPR*, 1–8 (2007).



Feng Li received his BE degree from the Department of Electrical Engineering, Fuzhou University, in 2003, and the MS degree from the Institute of Pattern Recognition, Shanghai Jiaotong University, in 2006. He is now a PhD student at the Department of Computer and Information Sciences, University of Delaware. His research interests include computational photography and applications for multicamera arrays.

Jian Sun is a lead researcher in the Visual Computing Group at Microsoft Research Asia. He received BS, MS, and PhD degrees from Xian Jiaotong University in 1997, 2000, and 2003. His research interests are interactive computer vision, Internet computers vision, stereo matching, and computational photography.



Jue Wang is a senior researcher at Creative Technologies Laboratory Adobe Systems Incorporated. He received his BE and MS from the Department of Automation, Tsinghua University, Beijing, China, in 2000 and 2002, and his PhD from the Department of Electrical Engineering, University of Washington in Seattle, in 2007. His research interests include computational photography and video, image and video processing, advanced computer graphics and vision techniques for better user experiences, online graphics, and vision applications.

Jingyi Yu is an assistant professor in the computer and information science department at the University of Delaware. He received his BS from Caltech in 2000, and MS and PhD degrees in EECS from MIT in 2005. His research interests span a range of topics in computer graphics, computer vision, and image processing, including computational photography, medical imaging, nonconventional optics and camera design, tracking and surveillance, and graphics hardware.