# Depth Estimation and Image Restoration Using Defocused Stereo Pairs

A.N. Rajagopalan,
S. Chaudhuri, *Senior Member*, *IEEE*, and
Uma Mudenagudi

**Abstract**—We propose a method for estimating depth from images captured with a real aperture camera by fusing defocus and stereo cues. The idea is to use stereo-based constraints in conjunction with defocusing to obtain improved estimates of depth over those of stereo or defocus alone. The depth map as well as the original image of the scene are modeled as Markov random fields with a smoothness prior, and their estimates are obtained by minimizing a suitable energy function using simulated annealing. The main advantage of the proposed method, despite being computationally less efficient than the standard stereo or DFD method, is simultaneous recovery of depth as well as space-variant restoration of the original focused image of the scene.

**Index Terms**—Defocus, stereo, disparity, Markov random field, blur identification, depth recovery.

## 1 Introduction

An important area of research in computer vision is the recovery of 3D information of a scene from 2D images. In humans, stereoscopically presented images provide information about depth. Julesz [1] showed that random dot stereograms provide a cue for disparity even when each image does not provide any high-level cue for depth. Interestingly, Pentland [2] reported a finding that the gradient of focus inherent in biological systems is also a useful source of depth information.

Binocular stereo matching is, in general, ambiguous if the matching is evaluated independently at each point purely by using image properties. All stereo matching algorithms examine candidate matches by calculating how much support they receive from their local neighborhood. Marr and Poggio [3] proposed a cooperative stereo algorithm based on a multiresolution framework. Barnard and Thompson [4] proposed a feature-based iterative algorithm to solve the correspondence problem. A large number of papers have appeared in the literature on stereo analysis and a review of them can be found in [5]. Conventional stereo analysis assumes an ideal pin-hole camera model which offers an infinite depth of field. However, any practical camera system will produce depth-related blurring. In the depth from defocus (DFD) technique, two images of an object, which may or may not be focused, and acquired with different camera settings, are processed to determine depth. The relative blur between the defocused images serves as a cue for depth. Unlike stereo, DFD uses a real aperture camera model (which is more practical). Since the early works of Pentland [2], several approaches [6], [7], [8], [9], [10] have emerged for solving the DFD problem. A comparative analysis of DFD and stereo can be found in [11]. Related works on defocus blur estimation in conjunction with image motion/disparity are addressed in [12], [13], [14], [15].

It is well-known that stereo yields accurate estimates of depth but the depth map is sparse since correspondence can be obtained with confidence only at prominent feature points. On the other hand, DFD gives a dense depth map but the accuracy of depth estimates are generally inferior to that of stereo-based methods. Estimates of depth from defocus can be particularly poor in severely blurred regions [8]. In this paper, we relax the assumption of a pin-hole camera model and propose an algorithm to recover depth from defocused stereo (DFDS) pairs of images. Defocus and stereo cues are fused to obtain improved dense estimates of depth. The additional constraints provided by stereo help in refining the estimates of depth over those obtained using DFD alone. Tsai et al. [16] have also proposed a scheme for integrating stereo and defocus. But, they use estimates from DFD to only initialize their stereo matching algorithm.

In stereo, the disparity is directly related to depth, while in DFD, it is the blur parameter that relates to depth. Hence, disparity can be expressed in terms of the blur parameter, the lens settings, and the base-line distance. This information is used to fuse the two methods, thereby deriving the advantages of both the methods. In DFD, since the blur depends on the depth of the scene, the point spread function (PSF) in turn becomes a function of depth. For simplicity, most techniques assume local space-invariance and compute depth. However, this can lead to poor estimates of depth due to the image overlap problem [9]. We model the depth and the focused image of the scene individually as Markov random fields (MRFs). Our approach avoids windowing and addresses the DFDS problem in its generality. Given two defocused stereo pairs of images, we obtain a focused image of the scene and a dense depth (blur/disparity) map. An important advantage of our method is that it not only recovers estimates of depth but also performs effective space-variant image restoration. The performance of the method is validated on synthetic as well as real images. The accuracy of depth estimates is superior compared to those obtained from DFD or the stereo method. A dense depth map is estimated without correspondence and interpolation. The quality of the restored image is also quite good.

Section 2 describes the framework for fusing defocus and stereo. In Section 3, we discuss the proposed approach for solving the DFDS problem. Experimental results are given in Section 4 while Section 5 concludes the paper.

## 2 Fusion of Defocus and Stereo

The basic structure of our scheme is given in Fig. 1. Note that we have two stereo pairs of images. The blurring of the two stereo pairs are different and so is their disparity since the image pairs are captured with two different camera settings. We attempt to simultaneously estimate blur (disparity) and restore one of the focused images of the scene in the defocused stereo pairs (say the left image). Estimating the other stereo pair is trivial once we know the disparity. As in most literature on stereo, we assume epipolar line constraint so that the disparity is only along one direction, say the $j$-direction. For the given observation model, the right image is given by

$$x_R(i,j) = x_L(i, j + d(i,j)) + w(i,j), \tag{1}$$

where $d(i,j)$ is the disparity associated with the stereo pair at location $(i,j)$. The noise $w(i,j)$ is assumed to be zero-mean and white Gaussian.

Because we use a practical real aperture camera model, points not in focus will appear blurred and the blur parameter $\sigma_k$ for the $k$th lens setting is given by

$$\sigma_k = \rho\, r_k V_k \left(\frac{1}{F_k} - \frac{1}{V_k} - \frac{1}{D}\right),\ k = 1, 2, \tag{2}$$

where $\rho$ is a camera constant, $r_k$ is the radius of the lens aperture, $F_k$ is the focal length, and $V_k$ is the image plane-to-lens distance for the chosen lens-setting [8]. Given two defocused images of a scene captured with different sets of camera parameters, the blur parameter at location $(i,j)$ for the two defocused images can be shown to be related by

$$\sigma_1(i,j) = \alpha\, \sigma_2(i,j) + \beta, \tag{3}$$

- *A.N. Rajagopalan is with the Department of Electrical Engineering, Indian Institute of Technology, Chennai 600 036, India. E-mail: raju@ee.iitm.ernet.in.*
- *S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology, Mumbai 400 076, India. E-mail: sc@ee.iitb.ac.in.*
- *U. Mudenagudi is with the Department of Electrical Engineering, B.V.B College of Engineering, Hubli 580 031, India. E-mail: umakm@yahoo.com.*

*Manuscript received 4 Sept. 2002; revised 29 Oct. 2003; accepted 8 Apr. 2004. Recommended for acceptance by R. Kumar.*
*For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117241.*

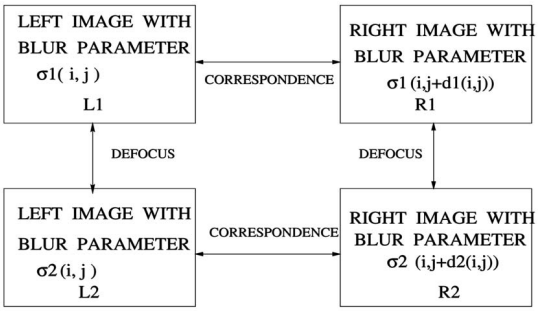


Fig. 1. Basic structure of depth from defocused stereo (DFDS).

where

$$\alpha = \frac{r_1 V_1}{r_2 V_2} \text{ and } \beta = \rho r_1 V_1 \left( \frac{1}{F_1} - \frac{1}{V_1} - \frac{1}{F_2} + \frac{1}{V_2} \right).$$

Thus, $\alpha$ and $\beta$ are known constants that depend on the camera settings. Most DFD methods assume a Gaussian-shaped blur model for the camera PSF, i.e., $h(i,j) = \frac{1}{2\sigma^2} \exp(-\frac{i^2+j^2}{2\sigma^2})$. Though the Gaussian blur is of infinite extent, a finite spatial extent approximation ($\pm 3\sigma$ pixels) is reasonable to assume for the Gaussian window. Note that the PSF is space-varying since $\sigma$ (as given in (2)) depends on the depth of the scene for a fixed camera setting.

From standard stereo analysis, we know that the depth ($D$) is related to the disparity ($d$), the baseline distance ($b$), and the focal length $f$ of the camera. If the focal length of the camera is changed, then for the same depth, the disparity changes. Let $d_k$ be the disparity and $f_k$ be the focal length associated with the image with blur parameter $\sigma_k$, $k = 1, 2$. For stereo [3], we can write

$$d_k = \frac{b f_k}{D}, \quad k = 1, 2. \tag{4}$$

Eliminating $D$, we get

$$d_1 = \frac{f_1}{f_2} d_2.$$

If we now relax the pin-hole camera model for stereo and substitute the value of depth in terms of disparity, we obtain disparity as a function of blur parameter and camera settings, i.e.,

$$d_k = b f_k \left( \frac{1}{F_k} - \frac{1}{V_k} - \frac{\sigma_k}{\rho r_k V_k} \right), \quad k = 1, 2.$$

If we assume $f_k = V_k$ (since the focal length $f_k$ in a pin-hole model is the same as $V_k$ in the DFD system), the above equation reduces to

$$d_k = b \left( \frac{V_k}{F_k} - \frac{\sigma_k}{\rho\, r_k} - 1 \right), \quad k = 1, 2. \tag{5}$$

From the above analysis, once the blur is estimated, the disparity can be calculated from the known camera settings. Thus, we can get a dense depth map without explicitly solving the correspondence problem.

## 3 Depth Recovery and Image Restoration

In Section 2, we described the geometric relation governing defocus and stereo. In the intensity domain, the relation between the focused and defocused stereo pairs is given by the following observation models:

$$\begin{aligned} g_{L_k} &= H_k x_L + w_{L_k} \\ g_{R_k} &= H_k(d_k) x_R + w_{R_k}, \ k = 1, 2, \end{aligned} \tag{6}$$

where the image and noise values have been lexicographically ordered. The noise terms $w_{L_k}$ and $w_{R_k}$ are assumed to be independent, white Gaussian with zero-mean and variance $\sigma_w^2$. The vectors $g_{L_k}$ and $g_{R_k}$ represent the observed defocused images with blur parameter $\sigma_k$ for the $k$th left and right stereo images, respectively. The blur matrix $H_k$ corresponds to the space-variant blurring function

$$h_k(i,j;m,n) = \frac{1}{2\pi\sigma_k^2(m,n)} \exp\left( -\frac{(i-m)^2 + (j-n)^2}{2\sigma_k^2(m,n)} \right).$$

Note that $H_k(d_k)$ is the same as $H_k$ with a shift due to disparity. The relation as expressed in (6) matches the left and the right images at the correct location (disparity) and for the correct amount of blurring which is also a function of disparity. By using a single parameter disparity, we have eliminated the (dependent) blur parameter. Since we do not assume local space-invariance, the blur parameter changes with the spatial location. Hence, the matrix $H_k$ will not be doubly block-Toeplitz. Given the observation models in (6) and the relations in (5), we attempt to solve for the estimates of depth (blur/disparity) and the focused image of the scene. Note that one needs to estimate only $x_L$ (or $x_R$) since they are just shifted versions of one another. From (3), it is clear that we need to estimate either $\sigma_1(i,j)$ or $\sigma_2(i,j)$, $\forall\, i, j$. If $\sigma_k$ is known, the disparity can be calculated from (5).

The problem of recovering the focused image and the space-variant blur parameter given two defocused stereo pairs is ill-posed and may not yield a unique solution, unless additional constraints like smoothness are added to restrict the solution space. We model both the space-variant blur parameter and the focused image of the scene as Markov random fields (MRFs). The concept of modeling depth/image as an MRF is well-known in the literature [17]. Since the change in the depth of a scene is usually gradual, the space-variant blur parameter also tends to have local dependencies. The utility of the MRF model lies in its ability to capture local dependencies and in its equivalence to the Gibbs random field [18]. Moreover, the MRF model preserves locality in the posterior distribution. This helps in reducing the computational complexity substantially.

Let $S$ denote the random field corresponding to the space-variant blur parameter $\sigma_1$ and $X_L$ denote the random field corresponding to the left focused image $x_L$ (the intensity process). Let $G_{L_1}$, $G_{L_2}$, $G_{R_1}$, and $G_{R_2}$ represent random fields corresponding to the four observed images, $g_{L_1}$, $g_{L_2}$, $g_{R_1}$, and $g_{R_2}$, respectively. The random field $S$ is assumed to be statistically independent of both $X_L$ and noise field $W_{L_k}$. This assumption may not always hold good [19]. Let $S$ take $P$ possible levels and $X_L$ take $M$ possible levels.

Since $S$ and $X_L$ are modeled as MRFs, we can write

$$P(S = s) = \frac{1}{Z^s} \exp[-U^s(s)], \tag{7}$$

$$P(X_L = x_L) = \frac{1}{Z^{x_L}} \exp[-U^{x_L}(x_L)]. \tag{8}$$

Note that $s$ is the same as the blur parameter $\sigma_1$. In (7), $s$ has been used for notational consistency.The terms $U^s(.)$ and $U^{x_L}(.)$ correspond to the energy functions associated with the space-variant blurring process in the left image and the intensity processes in the left image, respectively. Given a realization of $S$, i.e., the blur parameter $\sigma_1(i,j)$, $\forall\, i, j$, the blurring function $h_1(\cdot,\cdot)$ is known and, hence, the matrix $H_1$ is also known. Moreover, $h_2(\cdot,\cdot)$ and matrix $H_2$ can then be determined because $\sigma_2(i,j) = \alpha\, \sigma_1(i,j) + \beta$. Using (5), the disparity can be calculated.

Given the four observed images, the a posteriori conditional joint probability of $S$ and $X_L$ is given by

$$\frac{P\left(S=s, X_L=x_L \mid G_{L_1}=g_{L_1}, G_{L_2}=g_{L_2}, G_{R_1}=g_{R_1}, G_{R_2}=g_{R_2}\right) = P(S=s, X_L=x_L) P\left(G_{L_1}=g_{L_1}, G_{L_2}=g_{L_2}, G_{R_1}=g_{R_1}, G_{R_2}=g_{R_2} \mid S=s, X_L=x_L\right)}{P\left(G_{L_1}=g_{L_1}, G_{L_2}=g_{L_2}, G_{R_1}=g_{R_1}, G_{R_2}=g_{R_2}\right)}. \tag{9}$$

From Bayes rule, and the independence of $S$ and $X_L$, the problem of simultaneous estimation of blur and focused image can be posed as the following MAP problem:

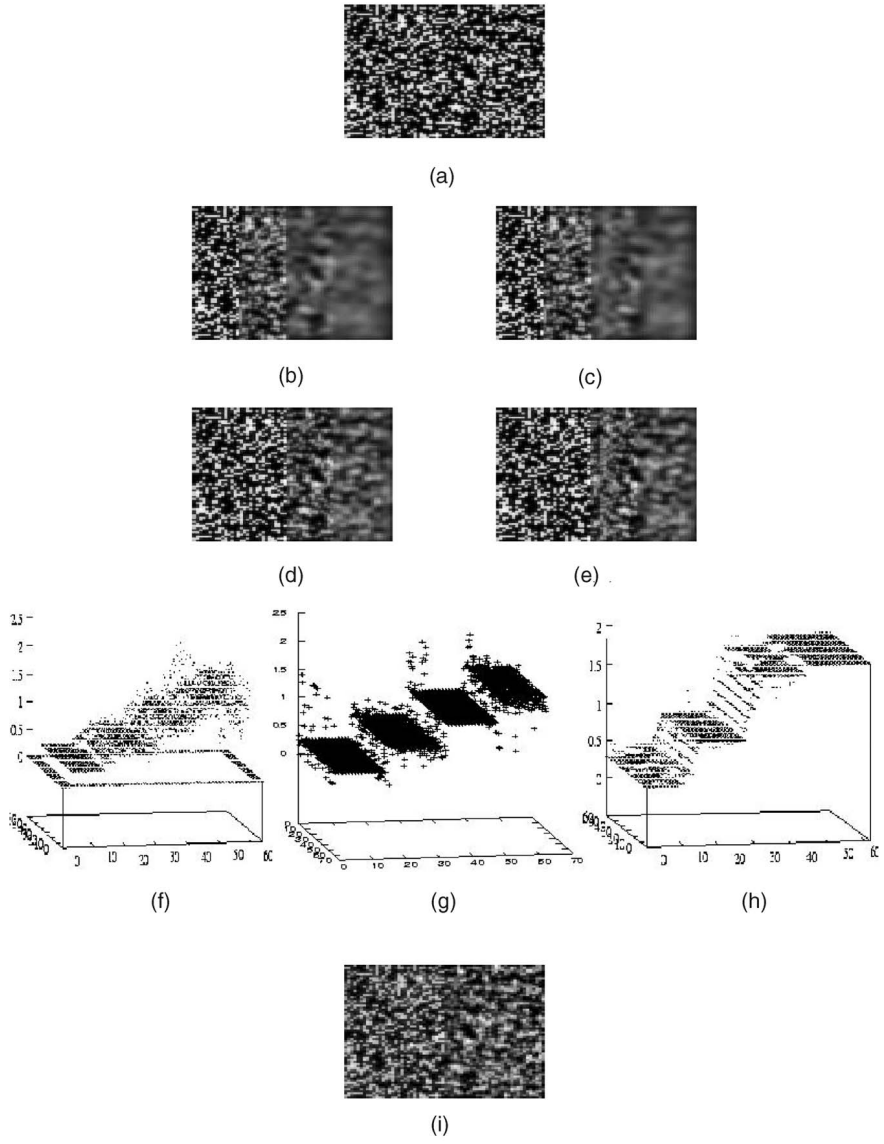


Fig. 2. (a) Original focused image. (b) Left defocused image with blur $\sigma_1$. (c) Stereo pair of (b). (d) Left defocused image with blur $\sigma_2$. (e) Stereo pair of (d). (f), (g), and (h) Values of $\sigma_1$ obtained using DFD alone, stereo alone, and the proposed method, respectively. (i) Space-variant restored image using DFDS.

$$\max_{s,x_L} P(G_{L_1} = g_{L_1}, G_{L_2} = g_{L_2}, G_{R_1} = g_{R_1}, G_{R_2} = g_{R_2} \mid S = s, X_L = x_L) P(S = s) P(X_L = x_L).$$

From (6) and using the fact that $S$ and $X_L$ are statistically independent of each other as well as the noise $W_{L_k}$, we get

$$-\log P(G_{L_1} = g_{L_1}, G_{L_2} = g_{L_2}, G_{R_1} = g_{R_1}, G_{R_2} = g_{R_2} \mid S = s, X_L = x_L) = \sum_{k=1}^{2} \frac{||g_{L_k} - H_k x_L||^2}{2\sigma_w^2} + \sum_{k=1}^{2} \frac{||g_{R_k} - H_k(d_k) x_R||^2}{2\sigma_w^2}. \quad (10)$$

Assuming first-order smoothness for the focused image as well as the blurring process, the posterior energy function to be minimized can be equivalently written as

$$U^P(s, l^s_{i,j}, v^s_{i,j}, x_L, l^{x_L}_{i,j}, v^{x_L}_{i,j}) = U_{def\_data} + U_{sm\_blur} + p_{sm\_blur} + U_{sm\_int} + p_{sm\_int} + U_{st\_data}, \quad (11)$$

where the horizontal and vertical binary line fields corresponding to the blurring process and the intensity image are denoted by $l^s_{i,j}$, $v^s_{i,j}$, $l^{x_L}_{i,j}$, and $v^{x_L}_{i,j}$, respectively. Line fields have been incorporated into the energy function in order to preserve the discontinuities [18].

In (11), the term $U_{def\_data}$ corresponds to data fitting based on the defocus information and is given (from (10)) as

$$U_{def\_data} = \frac{||g_{L_1} - H_1 x_L||^2}{2\sigma_w^2} + \frac{||g_{L_2} - H_2 x_L||^2}{2\sigma_w^2} + \frac{||g_{R_1} - H_1(d_1) x_R||^2}{2\sigma_w^2} + \frac{||g_{R_2} - H_2(d_2) x_R||^2}{2\sigma_w^2}. \quad (12)$$

The term $U_{sm\_blur}$ incorporates smoothness for blur and is given as

$$U_{sm\_blur} = \sum_{i,j} \lambda_s [(s_{i,j} - s_{i,j-1})^2 (1 - v^s_{i,j}) + (s_{i,j+1} - s_{i,j})^2 (1 - v^s_{i,j+1}) + (s_{i,j} - s_{i-1,j})^2 (1 - l^s_{i,j}) + (s_{i+1,j} - s_{i,j})^2 (1 - l^s_{i+1,j})], \quad (13)$$

where $\lambda_s$ is the regularization parameter corresponding to the blur parameter. Penalty term $p_{sm\_blur}$ is used to prevent spurious discontinuities and is given by

$$p_{sm\_blur} = \sum_{i,j} \gamma_s [l^s_{i,j} + l^s_{i+1,j} + v^s_{i,j} + v^s_{i,j+1}], \quad (14)$$

where $\gamma_s$ is the associated weight for penalty.

Along similar lines, we have a smoothness term $U_{sm\_int}$ and a penalty term $p_{sm\_int}$ for the intensity process also and these are given as follows:

$$U_{sm\_int} = \sum_{i,j} \lambda_x [(x_{L_{i,j}} - x_{L_{i,j-1}})^2 (1 - v_{i,j}^{x_L}) + (x_{L_{i,j+1}} - x_{L_{i,j}})^2 (1 - v_{i,j+1}^{x_L}) + (x_{L_{i,j}} - x_{L_{i-1,j}})^2 (1 - l_{i,j}^{x_L}) + (x_{L_{i+1,j}} - x_{L_{i,j}})^2 (1 - l_{i+1,j}^{x_L})]$$
$$\text{and} \quad p_{sm\_int} = \sum_{i,j} \gamma_x [l_{i,j}^{x_L} + l_{i+1,j}^{x_L} + v_{i,j}^{x_L} + v_{i,j+1}^{x_L}], \tag{15}$$

where $\lambda_x$ is the regularization parameter corresponding to the intensity image while $\gamma_x$ weights the penalty for introducing a discontinuity. Finally, the term $U_{st\_data}$ corresponds to data fitting due to stereo and is given as

$$U_{st\_data} = \lambda_{st} \left[ ||g_{R_1} - g_{L_1}(d_1)||^2 + ||g_{R_2} - g_{L_2}(d_2)||^2 \right], \tag{16}$$

where the parameter $\lambda_{st}$ weights how well the stereo image pairs match in terms of disparity.

The posterior energy function given by (11) is nonconvex and algorithms based on steepest-descent are prone to get trapped in local minima. We choose the simulated annealing (SA) algorithm [18] for minimizing the posterior energy function so as to obtain estimates of the space-variant blur parameter and the focused image simultaneously. For this purpose, a temperature variable is introduced in the objective function. The cooling schedule was chosen to be linear. Since the random fields associated with SV blur and image are assumed to be statistically independent, the values of blur $s_{i,j}$ and image $x_{i,j}$ at every location $(i, j)$ are changed independently. Parameter estimation in MRF is a difficult task. Here, we choose the MRF parameters in an ad hoc way.

From (11), it may be noted that the estimates of depth and the focused image of the scene using DFD alone can be obtained by simply leaving out the stereo terms [9]. However, as we shall show in the next section, the additional constraints provided by stereo are very useful in refining the estimates of depth.

## 4 Experimental Results

We demonstrate the performance of the proposed method in estimating blur (or disparity or depth) and restoring the focused image. The number of discrete levels for the blur parameter was chosen to be $64$. For the intensity process, $256$ levels were used (same as the CCD dynamic range). For DFD and DFDS, the window-based method of Subbarao [7] was used to obtain initial estimates of $\sigma_1$. The method of Roy and Cox [20] and the implementation available at http://www2.iro.umontreal.ca/~roys/publi/iccv98/code.html was used to get the stereo estimates.

In the first experiment, defocused versions of a random dot-pattern image (Fig. 2a) were generated. The blurring was stair-case type and we chose $\sigma_2(i, j) = 0.5\sigma_1(i, j)$. Figs. 2b, 2c, 2d, and 2e show the two defocused stereo pairs of images thus generated. Fig. 2f shows the initial estimates of the blur $\sigma_1$ using DFD alone (i.e., without the stereo constraints). Although one can make out the stair-case nature of the blur, the estimates are not very satisfactory. The $rms$ value of the error in the estimate of the blur was found to be $0.55$. Since the blur parameter and disparity are related, estimates of the blur obtained using focused left-right stereo pairs are shown in Fig. 2g. The $rms$ value of the error is $0.38$. The proposed method was next used to estimate the blur/disparity and the original focused image. The values of the various parameters used in the SA algorithm were as follows: $T_0 = 10.0$, $\lambda_s = 5000.0$, $\lambda_f = 0.005$, $\lambda_{st} = 0.01$, $\gamma_s = 10.0$, $\gamma_f = 15.0$, $\theta_s = 0.4$, $\theta_f = 25.0$, $\sigma_s = 0.1$, $\sigma_f = 6.0$, number of annealing iterations $= 200$, and the number of metropolis iterations $= 100$. Here, $T_0$ is the initial temperature, while $\theta_s$ and $\theta_f$ are thresholds for deciding the presence of an edge in the blur and in the image, respectively. The variances ${\sigma_s}^2$ and ${\sigma_f}^2$ are used in a Gaussian sampler to generate new samples of blur and intensity values and these are then used in the SA algorithm to generate a new realization. The estimated SV blur parameter and the restored image using the proposed method are shown in Figs. 2h and 2i, respectively. From the figure, it can be seen that the blur is well-captured and the edges are sharper. The improvement is particularly significant in regions of large blur where defocus is known to perform poorly due to reduced spectral content. This is also reflected in the $rms$ error which now reduces to $0.12$. Also, the restored image (which is an estimate of the original image in Fig. 2a) is quite good.

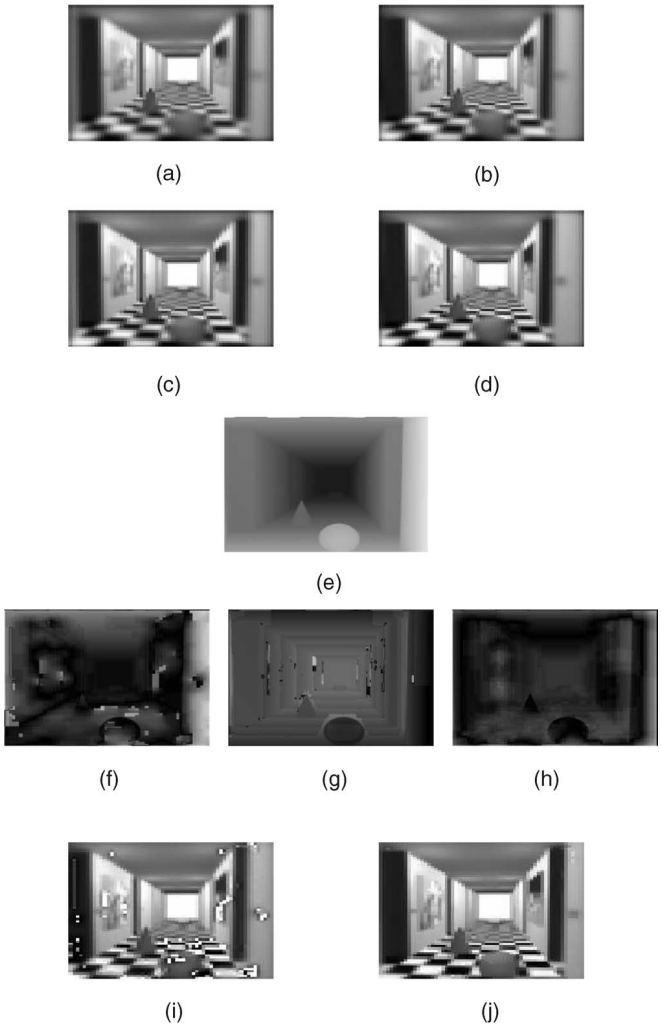


Fig. 3. (a) Left defocused image. (b) Right defocused stereo image with the same camera settings as in (a). (c) Left defocused image with different camera parameters. (d) Right defocused stereo pair with same camera settings as in (c). (e) True values of depth. (f) Error in the estimated values of depth using the DFD scheme. (g) Error in the estimated values of depth using only stereo. (h) Error in the estimated values of depth using the proposed scheme. (i) Reconstructed image using DFD alone. (j) Reconstructed image using defocus as well as stereo cues.

The method was next tested on the corridor image (obtained from CIL/CMU database) and the defocused stereo image pairs are given in Figs. 3a, 3b, 3c, and 3d. The defocused images were obtained from the ground-truthed disparity and depth values in conjunction with an appropriate choice of camera parameters. The true depth values are given in Fig. 3e. The error in the estimates of depth using the DFD method is shown in Fig. 3f. A darker gray level implies smaller error. The restored image is shown in Fig. 3i. The normalized $rms$ error for the depth map using DFD is $0.209$. We note that the estimates are poor at places of large blur where there is not enough spectral content. Note that the ball is quite severely blurred and, hence, the depth estimates in that region are not good for the DFD method. Depth estimates were next obtained using focused stereo pairs of the corridor image. The error in the estimates of depth

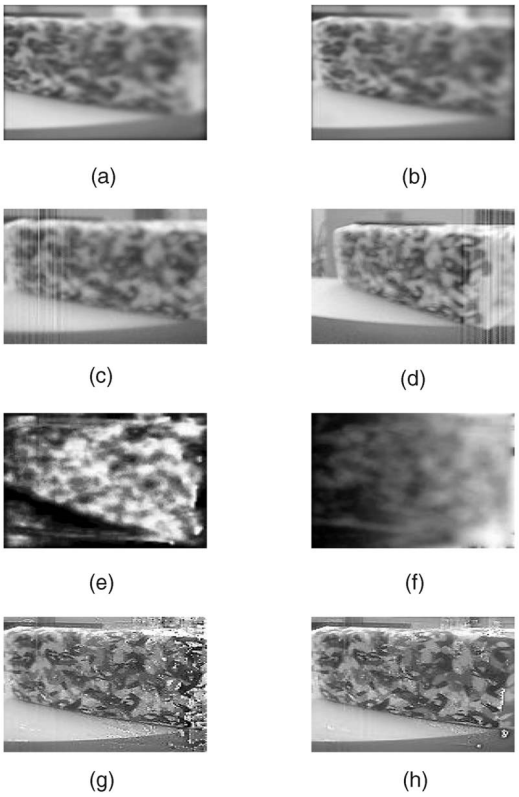


Fig. 4. (a) Left defocused image. (b) Stereo pair of (a). (c) Left defocused image with different camera settings. (d) Stereo pair of (c). (e) Estimated values of depth using only DFD. (f) Estimated values of depth using the proposed scheme. (g) Reconstructed image using DFD. (h) Restored focused image of the scene using the proposed method.

are displayed in Fig. 3g and the normalized $rms$ error is $0.264$. Results corresponding to the proposed scheme are shown in Figs. 3h and 3j. By fusing defocus and stereo cues, the depth estimates clearly improve. The error values are smaller compared to both DFD and stereo. We note that the estimates of depth near the cone and the ball are quite good. The $rms$ value of the error in the estimate of depth reduces to $0.149$. Because the depth estimates for DFDS are better, the restored image is also cleaner with few artifacts. The ball is quite focused and emerges nicely in the restored image (Fig. 3j) using DFDS as compared to DFD (Fig. 3i).

Finally, the performance of the proposed scheme was tested on images captured in our laboratory. The left and right defocused stereo pairs are shown in Figs. 4a, 4b, 4c, and 4d. Using DFD alone, the depth map and the focused image were first estimated and these are shown in Figs. 4e and 4g, respectively. From the figures, we note that although the recovered image is reasonably good, the depth map is not satisfactory. The estimates of depth and the focused image using the proposed scheme are given in Figs. 4f and 4h, respectively. The improvement due to fusion of defocus and stereo is amply evident. Some of the visible artifacts that were present in the restored image using DFD alone are now completely gone (particularly at the end nearer to the camera). The restored image is uniformly focused everywhere and is of very good quality. Also, the planar nature of the variation in the depth of the scene is better brought out by the proposed method. Since a real aperture camera cannot bring all points in a 3D scene simultaneously into focus, we do not have focused image pairs of the scene for computing depth using stereo alone. A major significance of our work lies in the following fact: Due to the physics of the problem, no real aperture camera can yield the image that the proposed method has been able to produce in Fig. 4h. In effect, Fig. 4h is a synthesis of the focused image of the scene had the camera brought the entire 3D scene into focus (this is not practically possible since a real aperture camera can only bring points at a single depth into focus).

## 5 Conclusions

We have proposed a new method for estimating depth that combines defocus and stereo cues for images captured with a real aperture camera. The method uses ideas from both DFD and stereo to its advantage. The estimates of depth are superior compared to both DFD and stereo. The recovered depth map is dense and no feature matching or interpolation is required. The method also simultaneously restores a focused image of the scene.

## Acknowledgments

S. Chaudhuri gratefully acknowledges the partial support received under the Swarnajayanti Fellowship scheme.

## References

[1] B. Julesz, "Binocular Depth Perception without Familiarity Cues," *Science,* vol. 145, pp. 356-362, 1964.
[2] A.P. Pentland, "A New Sense for Depth of Field, *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, pp. 523-531, 1987.
[3] D. Marr and T. Poggio, "Co-Operative Computation of Stereo Disparity," *Science,* vol. 194, pp. 283-287, 1976.
[4] S.T. Barnard and W.B. Thompson, "Disparity Analysis of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, pp. 333-339, 1980.
[5] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision,* vol. 47, pp. 7-42, 2002.
[6] M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 149-155, 1988.
[7] Y. Xiong and S.A. Shafer, "Depth from Focusing and Defocusing," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 68-73, 1993.
[8] A.N. Rajagopalan and S. Chaudhuri, "Space-Variant Approaches to Recovery of Depth from Defocused Images," *Computer Vision and Image Understanding,* vol. 68, pp. 309-329, 1997.
[9] A.N. Rajagopalan and S. Chaudhuri, "Optimal Recovery of Depth from Defocused Images Using an MRF Model," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1047-1052, 1998.
[10] M. Watanabe and S.K. Nayar, "Rational Filters for Passive Depth from Defocus," *Int'l J. Computer Vision,* vol. 27, pp. 203-225, 1998.
[11] Y.Y. Schechner and N. Kiryati, "Depth from Defocus vs. Stereo: How Different Really Are They?" *Int'l J. Computer Vision,* vol. 39, pp. 141-162, 2000.
[12] J. Flusser and T. Suk, "Degraded Image Analysis: An Invariant Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, pp. 590-603, 1998.
[13] A. Kubota, K. Kodama, and K. Aizawa, "Registration and Blur Estimation Methods for Multiple Differently Focused Images," *Proc. IEEE Int'l Conf. Image Processing,* vol. 2, pp. 447-451, 1999.
[14] J. Flusser and B. Zitova, "Combined Invariants to Linear Filtering and Rotation," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 13, pp. 1123-1136, 1999.
[15] F. Deschenes, P. Fuchs, and D. Zioul, "Homotopy-Based Estimation of Depth Cues in Spatial Domain," *Proc. IEEE Int'l Conf. Pattern Recognition,* vol. 3, pp. 627-630, 2002.
[16] Y.P. Tsai, J. Leu, and C.H. Chen, "Depth Estimation by Integration of Depth-from-Defocus and Stereo Vision," Inst. of Information Science, Taipei, 1998.
[17] J. Subrahmonia, Y.P. Hung, and D.B. Cooper, "Model-Based Segmentation and Estimation of 3D Surfaces from Two or More Intensity Images Using Markov Random Fields," *Proc. IEEE Int'l Conf. Pattern Recognition,* pp. 390-397, 1990.
[18] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Distribution of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, pp. 721-741, 1984.
[19] D. Rajan and S. Chaudhuri, "Simultaneous Estimation of Super-Resolved Scene and depth Map from Low Resolution Observations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, pp. 1102-1117, 2003.
[20] S. Roy and I.J. Cox, "A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 492-499, 1998.