Blur-aware Disparity Estimation from Defocus Stereo Images

Ching-Hui Chen*, Hui Zhou†, and Timo Ahonen†
*Department of Electrical and Computer Engineering,
University of Maryland, College Park, MD, USA
†Nokia Technologies, Sunnyvale, CA, USA
ching@umiacs.umd.edu, hui.7.zhou@nokia.com, timo.ahonen@nokia.com

Abstract

Defocus blur usually causes performance degradation in establishing the visual correspondence between stereo images. We propose a blur-aware disparity estimation method that is robust to the mismatch of focus in stereo images. The relative blur resulting from the mismatch of focus between stereo images is approximated as the difference of the square diameters of the blur kernels. Based on the defocus and stereo model, we propose the relative blur versus disparity (RBD) model that characterizes the relative blur as a second-order polynomial function of disparity. Our method alternates between RBD model update and disparity update in each iteration. The RBD model in return refines the disparity estimation by updating the matching cost and aggregation weight to compensate the mismatch of focus. Experiments using both synthesized and real datasets demonstrate the effectiveness of our proposed algorithm.

1. Introduction

Recovery of the depth information from binocular vision is an essential task since the depth information is useful in the reconstruction of 3D shape [12], matting [24, 14], and generating variable focus images/videos [13, 22]. However, the performance of stereo correspondence degrades in regions without prominent textures. Even with the support of the local neighborhood and regularization [2, 27], stereo matching remains a challenging problem.

Most of the stereo matching techniques [19] assume stereo images are all-in-focus. Nevertheless, such assumption is not always satisfied since cameras with real aperture cannot provide infinite depth of field. For instance, mobile devices usually equip a camera of small f-number (wide aperture relative to focal length) that is capable of gathering much light but limits the depth of field. Hence, the visual correspondence becomes weak when two corresponding regions experience unequal amount of defocus blur. While establishing the visual correspondence between stereo images captured with inconsistent focus settings, we cannot simply ignore the effect of defocus blur. Although the mismatch of focus is usually adverse in stereo matching, the relative blur between stereo images provides additional cue to resolve the intrinsic ambiguity of regular stereo matching, such as establishing correspondence in region of repetitive texture. Our objective is to improve disparity estimation for defocus stereo images via compensating the mismatch of focus and integrating the cue of relative blur as a whole.

Defocus can be complementary cue to resolve the inherent ambiguity of stereo matching [10, 20]. Several approaches have fused the information from disparity and defocus blur for stereo matching. Rajagopalan *et al.* [18] propose a Markov random field (MRF) based approach to utilize depth from defocus [6] and stereo matching for robust depth estimation. Their proposed approach operates on two pairs of stereo images, where each view possesses a focal stack of two images. Assuming that the camera parameters and baseline are known, the depth estimation is modeled as an energy minimization framework, where the mismatch of stereo correspondence and deviation from the defocus model will be penalized. Therefore, it is more robust than using either regular stereo matching or depth from defocus alone. However, this method requires two focal stacks from each view, and the camera parameters should be known in advance. This method is generalized in [1] to couple the blur, motion, and depth using a calibrated setup. Several techniques estimate the point spread function without calibrating the camera parameters for depth inference [15, 21].

Takeda *et al.* [25] use a pair of defocus stereo images to combine the depth from defocus and stereo matching. This approach requires each lens equipped with a coded aperture mask, and the relation between disparity and the diameter of the blur kernel should be calibrated in advance. Although their proposed method is advantageous in recovering the all-in-focus image, it is not general for defocus stereo images captured by regular cameras with unrestricted focus

settings. Devernay *et al.* [3] propose an approach for detecting focus mismatch between views in stereoscopic content via seeking for a focus mismatch pattern among a set of legit patterns. This enables the real-time detection of mismatch of focus to provide feedback to the camera operator.

Since the camera parameters are not usually available when the stereo images are captured by point-and-shoot cameras, Li *et al.* [13] propose a disparity estimation method using a single pair of defocus stereo images. They present an iterative MRF-based method that refines the disparity by modeling the matching cost corresponding to defocus blur. Their method relies on estimating the camera parameters from the pixels with prominent textures in the in-focus plane of each view.

Our work is closely related to Li *et al.*'s method [13]. We propose the **relative blur versus disparity** (RBD) model, which characterizes relative blur as a second-order polynomial function of disparity. Note that the relative blur is approximated as the difference of the square diameters of the blur kernels, resulting from the mismatch of focus between stereo images. Most of the pixels with prominent textures are utilized to construct RBD model via curve fitting across all the disparity levels. This is more robust than only using the in-focus pixels to recover the camera parameters in prior art [13], where erroneous detection of in-focus pixels in the complex scenes can lead to unreliable estimation of camera parameters. Furthermore, Li *et al.*'s method relies on the assumption that the apertures of the two cameras are identical and the existence of in-focus plane in each view. While such assumption is valid under some circumstances, it may limit its applicability in the disparity estimation for defocus stereo images.

In contrast to prior works that require either the calibration of the focus related parameters or multiple images to form a focal stack in each view [25, 18], our method handles the general scenario of estimating the disparity from a pair of defocus stereo images. Camera parameters that affect defocus blur, including focus setting, the diameter of the aperture, are succinctly represented by the coefficients of RBD model to characterize the relative blur as a function of disparity. We adopt the non-local cost aggregation method for stereo matching [26] to illustrate the integration with our RBD model. Since the effect of mismatch of focus is compensated via updating the volume cost and the weight for cost aggregation, our proposed framework provides reliable estimation of disparity. Experiments using both synthesized and real data confirm the improvement of the accuracy and robustness of the proposed algorithm.

2. Model of Depth from Defocus Stereo

In the scenario of depth from defocus stereo, both spatially-variant blur and disparity provide the inference for depth information. To utilize all the cues, establishing the visual correspondence across two images should take disparity and blur into account.

In regular stereo matching, two all-in-focus stereo images are processed to determine disparity. The left image I_1 and right image I_2 are related with the spatially-variant disparity $\delta(p)$, and thus the correspondence between the two images can be modeled as

$$I_1(p) = I_2(p + \delta(p)), \tag{1}$$

where p is the spatial index of a pixel.

Considering the effect of defocus blur, we model the effect of spatially-variant blur as the convolution of the all-in-focus image and the blur kernel. Hence, the defocus stereo images, $\tilde{I}_1$ and $\tilde{I}_2$, are modeled by convolving the all-in-focus image pair I_1 and I_2 with spatially-variant blur kernel b_1 and b_2, respectively. We can model $\tilde{I}_1$ and $\tilde{I}_2$ as

$$\tilde{I}_1(p) = b_1(p) \otimes I_1(p) \;\text{ and }\; \tilde{I}_2(p) = b_2(p) \otimes I_2(p), \tag{2}$$

respectively. Note that the diameter of the blur kernel $b_k(p)$ is represented as $\sigma_k(p)$. Empirically, modeling the defocus blur with disk kernel is comparable to that with Gaussian kernel in our implementation. Similar to [5], we adopt the disk kernel as the blur kernel since it has finite support weight as compared to Gaussian kernel. Since each pair of pixels associated by the disparity can bear different amount of defocus blur in each view, associating the defocus stereo images by trivially using the original pixel can lead to severe performance degradation. Alternatively, we associate the defocus stereo images with their equally-defocused images by applying additional blur to the relatively in-focus regions. Hence, the correspondence can be modeled as

$$\begin{cases} b_r(p) \otimes \tilde{I}_1(p) \backsimeq \tilde{I}_2(p + \delta(p)), \text{ if } \sigma_2(p + \delta(p)) \geq \sigma_1(p). \\ \tilde{I}_1(p) \backsimeq b_r(p) \otimes \tilde{I}_2(p + \delta(p)), \text{ if } \sigma_2(p + \delta(p)) < \sigma_1(p). \end{cases} \tag{3}$$

For the ease of notation, the spatial indices for $b_r(p)$, $\delta(p)$ and $\sigma_k(p)$ are omitted henceforth. Note that b_r is the relative blur kernel applied to either one of the views such that both views are equally-defocused. The diameter of the relative blur kernel b_r can be approximated as

$$\sigma_r \backsimeq \sqrt{|\sigma_2^2 - \sigma_1^2|}. \tag{4}$$

Since the diameter of the relative blur kernel cannot indicate whether the left or right view is more in-focus than the other, we define the relative blur as

$$\Delta\sigma^2 \triangleq \sigma_2^2 - \sigma_1^2, \tag{5}$$

where $\Delta\sigma^2 > 0$ ($\Delta\sigma^2 < 0$) indicates a pixel in the left (right) view is more in-focus than its corresponding pixel in the right (left) view.

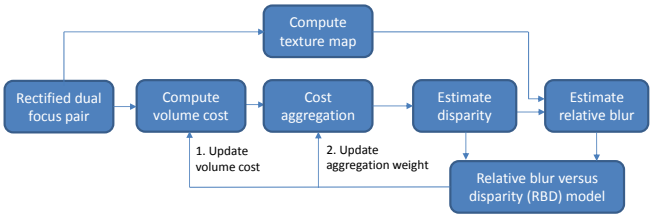


Figure 1: Block diagram of the proposed approach.

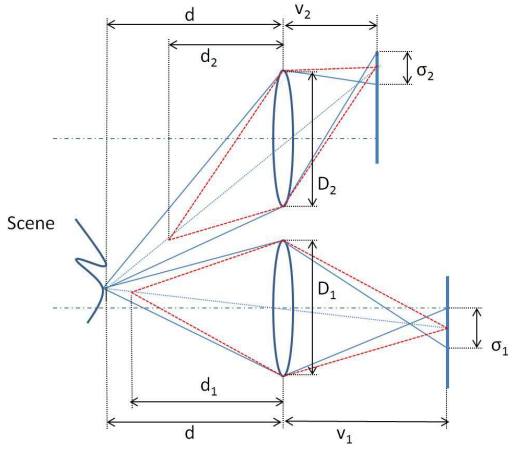


Figure 2: Optical geometry of defocus stereo based on thin lens model.

3. Disparity Estimation from Defocus Stereo

3.1. Overview of the Proposed Method

The block diagram of our proposed approach is shown in Figure 1. A pair of rectified defocus stereo images is the input for computing the initial matching cost, and we adopt the non-local cost aggregation method [26] to estimate the disparity. The defocus stereo pair is also used to generate the texture map, where pixels with prominent textures will be used to estimate the relative blur. The RBD model is estimated by fitting those samples carrying the information of relative blur and disparity. The estimated RBD model is then used to update the volume cost and aggregation weight. The entire process is repeated until the estimation of disparity converges or the maximum number of iterations is attained.

3.2. Relative Blur versus Disparity (RBD) Model

The depth from defocus and depth from stereo have been investigated in the literature to exploit the depth cues from focal stack and stereo images, respectively. In this paper, we assume the optical system of the camera obeys thin lens model [23, 9]. Given the focal length, diameter of the aperture, and focus setting, the defocus blur only depends on the depth of the scenes. The diameter of the blur kernel at distance d can be computed from the geometry of optical system [16, 8], as shown in Figure 2. It is represented as

$$\sigma_k = D_k \frac{|d - d_k|}{d} \frac{v_k}{d_k}, \tag{6}$$

where k indicates the parameter belonging to left view ($k = 1$) or right view ($k = 2$). D_k is the diameter of the aperture, and v_k is the focus setting defined as the distance between sensor plane and lens. d_k is the distance between in-focus plane and lens. From the thin lens model, we can interpret the focus setting as

$$v_k = \frac{d_k f_k}{d_k - f_k}, \tag{7}$$

where f_k represents the focal length.

Substituting (7) into (6), we reformulate (6) as

$$\sigma_k = D_k \frac{|d - d_k|}{d} \frac{f_k}{d_k - f_k}. \tag{8}$$

On the other hand, the depth can be computed from disparity. The depth corresponding to the disparity of in-focus plane δ_k and that of out-of-focus plane δ can be represented as

$$d_k = \frac{f_k B}{\delta_k} \quad \text{and} \quad d = \frac{f_k B}{\delta}, \tag{9}$$

respectively. Note that B is the baseline. We can replace d_k and d in (8) by (9), and (8) can be reformulated as

$$\sigma_k = D_k \frac{|\delta - \delta_k|}{B - \delta_k}. \tag{10}$$

Assuming the defocus stereo pair has been rectified and normalized to the same scale [7], both views have an equivalent focal length. Therefore, the measuring units for δ, δ_1, and δ_2 are identical.

According to (5) and (10), the RBD model can be formulated as

$$\begin{aligned}
&\Delta\sigma^2(\delta) = \sigma_2^2(\delta) - \sigma_1^2(\delta) \\
&= \left(D_2 \frac{|\delta - \delta_2|}{B - \delta_2}\right)^2 - \left(D_1 \frac{|\delta - \delta_1|}{B - \delta_1}\right)^2 \\
&= \left(\frac{D_2^2}{(B - \delta_2)^2} + \frac{-D_1^2}{(B - \delta_1)^2}\right) \delta^2 \\
&+ \left(\frac{-2D_2^2\delta_2}{(B - \delta_2)^2} + \frac{2D_1^2\delta_1}{(B - \delta_1)^2}\right) \delta + \left(\frac{D_2^2\delta_2^2}{(B - \delta_2)^2} + \frac{-D_1^2\delta_1^2}{(B - \delta_1)^2}\right) \\
&= X\delta^2 + Y\delta + Z.
\end{aligned} \tag{11}$$

It is clear that the focus settings, diameters of the apertures, and baseline, are constant parameters. Hence, the relative blur can be characterized by a second-order polynomial function of disparity without knowing the exact value of each parameter. Note that we directly model the relative blur versus disparity regardless of different aperture diameters and focal settings.

3.3. Estimating the Coefficients of RBD Model

In order to estimate the coefficients of RBD model, we assume that pixels belonging to the same disparity value possess identical diameter of the blur kernel. As a result, pairs of pixels with disparity value δ will have identical relative blur $\Delta\sigma^2(\delta)$. The coefficients of RBD model can be obtained by curve fitting using samples $(\delta, \Delta\sigma^2(\delta))$ collected from each disparity level.

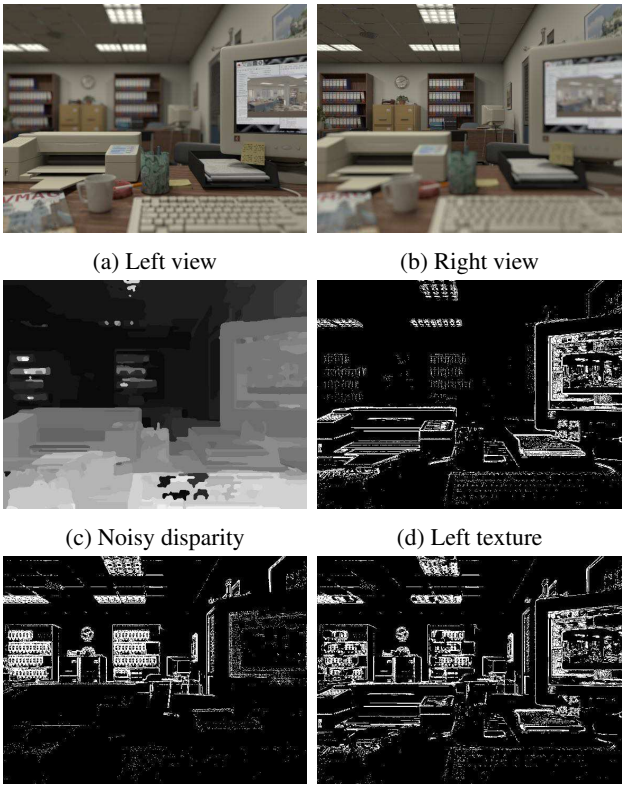
Given the disparity map from the initial estimation or previous iteration, we can compute the relative blur by searching for the diameter of the relative blur kernel b_r using (3). The estimation of relative blur becomes unstable in smooth region since homogeneous region undergoing different amount of defocus blur typically looks similar. Hence, we use edge detection methods (e.g., difference of Gaussians) to detect regions with prominent texture for estimating the relative blur. Since the texture region can be blurred by defocus in one view but well preserved in another view, we merge the textures of both views to utilize all the prominent textures. Given the disparity map of the right view, we map the edges detected from the right view to the left view to create the merged texture map. With the left and right disparity maps computed in the previous iteration, we can further refine the merged texture map by removing those edges that do not have consistent disparity between the left and right disparity maps. This step is similar to retain non-occluded pixels using cross-consistency check [4], but we tailor it to select the reliable textured pixels for the estimation of relative blur. Figure 3 shows the merged result of texture regions using the noisy initial disparity.

The disparity estimation in early iterations can be unstable due to mismatch of focus. In addition, pixel-wise relative blur estimation is prone to error in the boundaries of discontinuous depth. Hence, we propose an ensemble fusion scheme to reliably utilize the information of defocus stereo images. We estimate the relative blur from pixels belonging to the same disparity level to make a group decision. The estimated relative blur $\widehat{\Delta\sigma^2}(\delta)$ minimizes the mean square error of pixels belonging to the disparity level δ, which can be formulated as

$$\widehat{\Delta\sigma^2}(\delta) = \arg\min_{\Delta\sigma^2 \in \mathcal{K}} \mathrm{MSE}(\delta, \Delta\sigma^2), \tag{12}$$

where $\mathcal{K}$ is the set consisting of possible values of relative blur. $\mathrm{MSE}(\delta, \Delta\sigma^2)$ is the mean square error for stable pixels lying in disparity δ assuming relative blur $\Delta\sigma^2$ is applied, which is formulated as

$$\mathrm{MSE}(\delta, \Delta\sigma^2) = \begin{cases} \frac{1}{|\mathcal{L}(\delta)|} \sum_{p \in \mathcal{L}(\delta)} |b_r \otimes \tilde{I}_1(p) - \tilde{I}_2(p+\delta)|^2, \text{ if } \Delta\sigma^2 \geq 0, \\ \frac{1}{|\mathcal{L}(\delta)|} \sum_{p \in \mathcal{L}(\delta)} |\tilde{I}_1(p) - b_r \otimes \tilde{I}_2(p+\delta)|^2, \text{ if } \Delta\sigma^2 < 0, \end{cases} \tag{13}$$



(a) Left view (b) Right view (c) Noisy disparity (d) Left texture (e) Right texture (f) Merged texture

Figure 3: Textures from both views are merged using the initial disparity. The defocus stereo images in this illustration are rendered images from a 3D scene.

where $\mathcal{L}(\delta)$ is the set containing the spatial indices of stable pixels belonging to disparity δ, and $|\mathcal{L}(\delta)|$ returns the cardinality of $\mathcal{L}(\delta)$.

For each disparity level, we can obtain a sample $(\delta, \widehat{\Delta\sigma^2}(\delta))$ consisting of the relative blur and disparity value. Because the disparity map in early iterations is prone to error, and some of the disparity levels may have inadequate number of stable corresponding pairs, we propose to estimate the coefficients of (11) by weighted least squares fitting. That is, each sample $(\delta, \widehat{\Delta\sigma^2}(\delta))$ is weighted by

$$\mathrm{W}(\delta) = \frac{|\mathcal{L}(\delta)|}{\mathrm{MSE}(\delta, \widehat{\Delta\sigma^2}(\delta))}. \tag{14}$$

This implies that the disparity level that owns large number of stable pixels and is well modeled by (3) will receive a large weight when computing the coefficients of RBD model.

3.4. Blur-aware Non-local Cost Aggregation

We incorporate our RBD model with the non-local cost aggregation method proposed in [26] to handle the defocus

stereo images. Specifically, we formulate the blur-aware aggregation cost as

$$\tilde{C}_\delta^A(p) = \sum_q \exp(-\frac{\tilde{\mathcal{D}}(p,q)}{\alpha})\tilde{C}_\delta(q), \tag{15}$$

where $\tilde{\mathcal{D}}(p,q)$ is the distance between pixel p and pixel q in the minimum spanning tree (MST), and $\tilde{C}_\delta(q)$ is the cost of pixel q at disparity δ. Note that α is the constant that controls the similarity between two nodes in the MST. Both the volume cost and aggregation weight are updated using our estimated RBD model. The procedure of blur-aware cost aggregation can be explained in two folds. First, we formulate the blur-aware volume cost $\tilde{C}_\delta(q)$ as

$$\tilde{C}_\delta(q) = \begin{cases} C(b_r(\delta) \otimes \tilde{I}_1(q), \tilde{I}_2(q+\delta)), \text{ if } \Delta\sigma^2(\delta) \geq 0, \\ C(\tilde{I}_1(q), b_r(\delta) \otimes \tilde{I}_2(q+\delta)), \text{ if } \Delta\sigma^2(\delta) < 0, \end{cases} \tag{16}$$

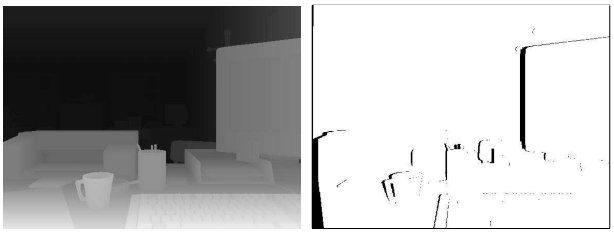
where the diameter of the relative blur kernel $b_r(\delta)$ is computed by $\sigma_r(\delta) = \sqrt{|\Delta\sigma^2(\delta)|}$, and $C(i,j)$ returns the data cost representing the dissimilarity between pixel i and j. Similar to [26], we use color and gradient as the features for computing data cost. The updated volume cost compensates the asymmetric defocus between stereo images by blur-aware disparity matching. Second, we update the aggregation weight by utilizing two views simultaneously. Although existing works have utilized both views for guidance, they typically assume information across two images are symmetric [11]. Nevertheless, this is not the case for the defocus stereo pairs, where both images can experience different defocus blur. Hence, we tailor the existing schemes to automatically select the guidance based on the hypothesized disparity with our RBD model. We compute the $\tilde{\mathcal{D}}(p,q)$ only from the much in-focus region in either left or right view, and thus the information in both images are utilized simultaneously.

4. Experimental Results

Since the dataset of defocus stereo images is not publicly available, we synthetically generate defocus stereo pairs from a 3D scene and its groundtruth for evaluation. In order to verify the effectiveness of our methods in real world, we collect defocus stereo images from cameras of real aperture for comparison.

4.1. Synthetic Datasets

We use the *average office* 3D scene provided by Jaime Vives Piqueres (http://www.ignorancia.org) to synthesize the defocus stereo images. In order to obtain the full control of camera parameters and settings, we synthesize the defocus stereo images of the scene from a ray-tracing tool [17] and generate its groundtruth disparity (Figure 4a) and non-occlusion mask (Figure 4b) as benchmark [3]. The disparity of synthetic stereo images varies from 0 to 24 pixels, and the disparity estimation is treated as a bad pixel if the estimation error is larger than one pixel. The technique of camera parameter estimation proposed by Li *et al.* [13] for defocus stereo matching is implemented for comparison.



(a) Groundtruth disparity (b) Non-occlusion mask

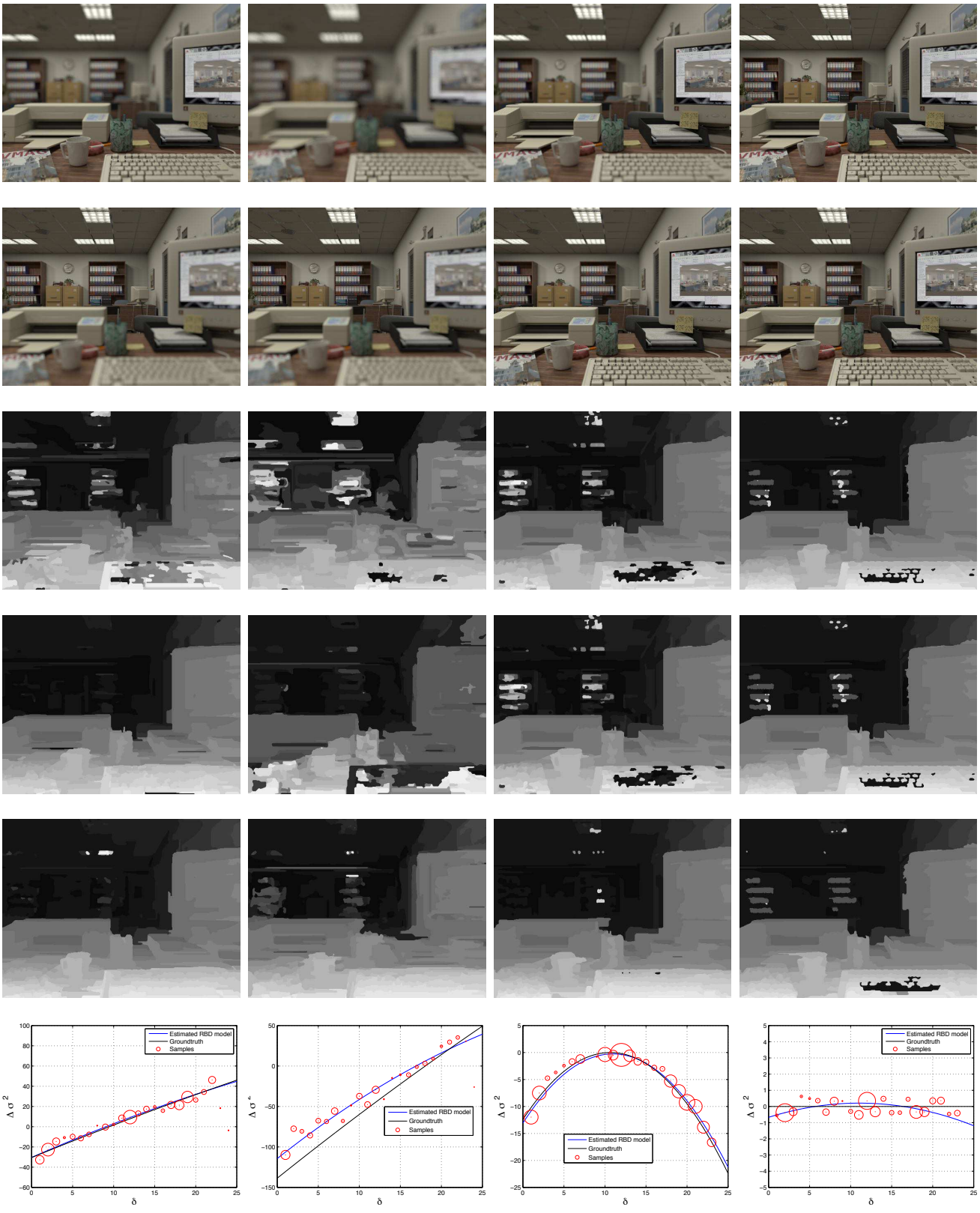
Figure 4: The groundtruth of the left view.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Yang [26]	36.36 / 35.92	57.14 / 56.49	17.56 / 17.33	9.74 / 9.05
Li *et al.* [13]	14.09 / 13.49	58.44 / 57.88	*17.56 / 17.33	*9.74 / 9.05
Ours	11.39 / 10.73	23.30 / 22.24	7.10 / 6.40	11.32 / 10.52

Table 1: Percentage of the bad pixels in the left view evaluated on synthetic data (all pixels / non-occlusion pixels).
* The results of Li *et al.* [13] for Dataset 3 and 4 are identical to that of Yang [26] since their model degenerates in the initial iteration.

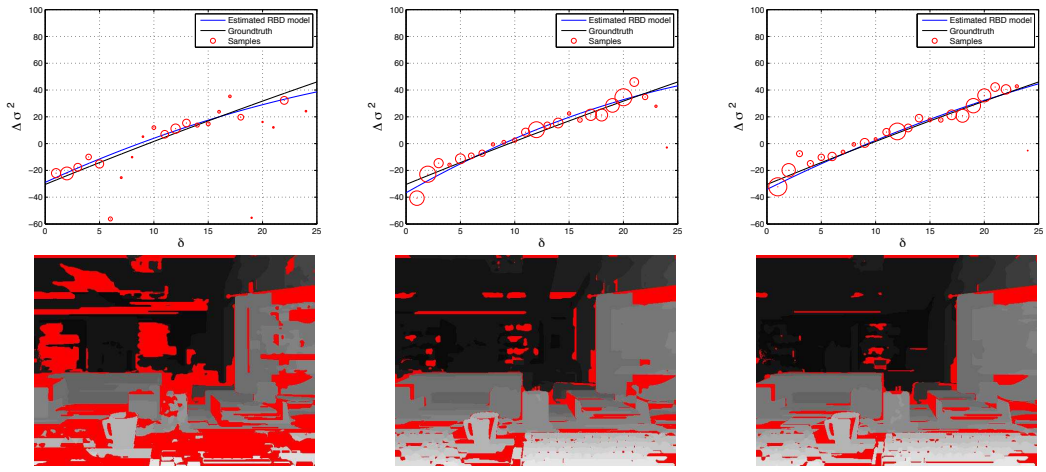
We generate four sets of defocus stereo images of various camera settings for evaluation. Each set of stereo images and their disparity estimation results are demonstrated in each column of Figure 5. The left and right views are shown in the first and second rows, respectively. Both views are with resolution 576×432 and f-number $f/2.2$. Throughout all the experiments in the synthetic dataset, we provide the result evaluated on the left view. The percentage of bad pixels for those erroneous disparity estimation are demonstrated in Table 1, and the disparity estimation results with annotated bad pixels are presented in the supplementary material.

In the first synthetic dataset, the left and right views focus on the near and far sites, respectively. Due to the asymmetric focus, the performance of disparity estimation based on [26] degrades since conventional stereo matching algorithm is not eligible to perceive the focal blur. Both of our method and Li *et al.* [13] successfully improve the disparity estimation. In terms of the quantitative performance on the percentage of bad pixels, our proposed method outperforms Li *et al.* since we can accurately estimate the coefficients of RBD model, which in return helps the disparity matching of defocus images by updating the volume cost. The estimation of RBD model is demonstrated in the last row of Figure 5, where the estimated RBD model mostly



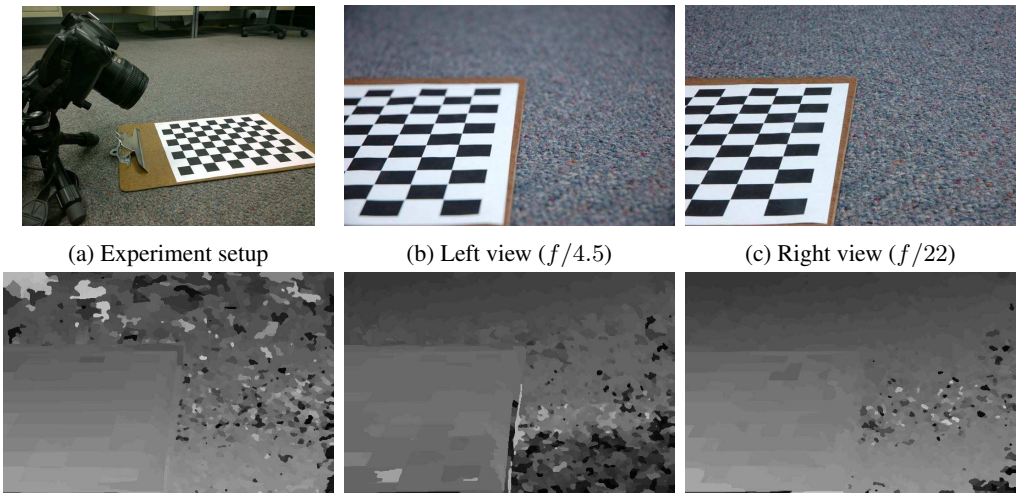
(a) Dataset 1 (b) Dataset 2 (c) Dataset 3 (d) Dataset 4

Figure 5: Performance evaluation on the synthetic dataset: Left views (the first row), right views (the second row), initial disparity estimation from [26] (the third row), disparity estimation of [13] (the fourth row), our proposed method (the fifth row), and the curve fitting of RBD samples (the last row). The disparity estimation results with annotated bad pixels are presented in the supplementary material.



(a) Iteration 1 (b) Iteration 3 (c) Iteration 5

Figure 6: The weighted least square fitting of RBD model at Iteration 1, 3, and 5, and the weight of the sample is proportional to the radius of the circle (the first row). The intermediate disparity map of the corresponding iteration, and the erroneous disparity values are marked in red (the second row).



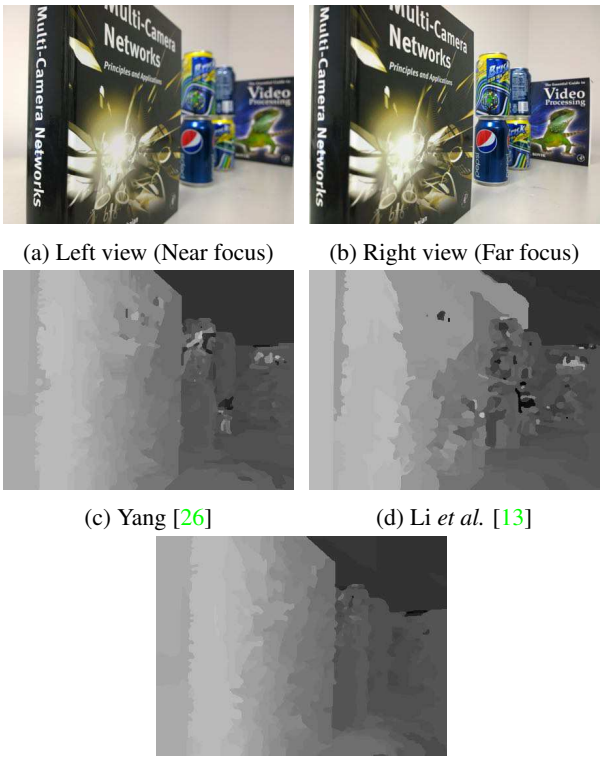
(a) Experiment setup (b) Left view ($f/4.5$) (c) Right view ($f/22$)

(d) Yang [26] (e) Li *et al.* [13] (f) Our proposed method

Figure 7: Experiment on repetitive texture in real world with asymmetric diameters of apertures.

coincides with the groundtruth computed from the actual camera parameters. Figure 6 demonstrates the intermediate results of the weighted least square fitting of RBD model using the first synthetic dataset. The estimated RBD model becomes closer to the groundtruth as the number of iterations increases. Also some of the outlier samples in the early iterations have been corrected in later process. Note that samples that deviate from the actual RBD model receive small weights (circles of small radius) as denoted by (14), and thus the impact of outlier is suppressed. Besides, we can observe the amount of bad pixels is reduced while the estimation of the RBD model is incrementally refined.

In the second synthetic dataset, the left camera focuses at the near site before the nearest object. Since the in-focus plane is not in the scene, the left view becomes fully defocused. The right view focuses on the far site, which is identical to the right view of the first synthetic dataset. It is clear that our proposed method outperforms Li *et al.* qualitatively and quantitatively since the method proposed by Li *et al.* assumes both views cover the in-focus planes. Our proposed method however do not rely on such assumption, so it can be applied to different types of defocus stereo images. In the third dataset, both views focus at the same depth. The depth of field is different since the left and right cameras



(a) Left view (Near focus) (b) Right view (Far focus)

(c) Yang [26] (d) Li *et al.* [13]

(e) Our proposed method

Figure 8: Experiment on a scene in real world with asymmetric focus settings and equal diameter of apertures ($f/2.2$).

have f-number $f/2.2$ and $f/8$, respectively. Our method correctly estimates the RBD model and improves the quality of disparity estimation. The method from Li *et al.* does not operate since their model becomes degenerated when both camera focus at the same depth.

In order to investigate the performance of our algorithm on nearly all-in-focus stereo images, we generate the fourth synthetic dataset with small f-number. Both cameras have f-number $f/22$ and focus at the same depth. Our method is slightly worse than Yang [26] as we intend to handle general defocus stereo images. It is interesting to notice that the percentage of bad pixels in the fourth dataset is slightly larger than that in the third dataset when we apply our method. Note that the disparity estimation in repetitive texture regions (e.g., the keyboard and the books on the bookshelves) is more erroneous in the nearly all-in-focus stereo images than in the third dataset using our method. This illustrates that our method can help resolve ambiguity of repetitive texture by jointly estimating the relative blur and disparity.

4.2. Real Datasets

We conduct a set of experiments on the repetitive pattern in real world to demonstrate that the ambiguity resulting from repetitive texture can be mitigated by integrating our RBD model. We use a Nikon D50 with Nikon AF-S DX Nikkor 18-70 mm lens to capture our test images. As our main objective is to verify the effectiveness of our proposed method, we assume the defocus stereo images are rectified or readily aligned while taken. The camera is mounted on a sliding bar, and it captures a scene with a checkerboard lying on the carpet (Figure 7a). The left and right views focus at the middle region of the scene. The focal length of lens is set to 70 mm. Images of both views are resized to resolution of 602×400. In this experiment, we use different apertures to create the asymmetric defocus effect. The left view (Figure 7b) and right view (Figure 7c) are captured with f-number $f/4.5$ and $f/22$, respectively. Note that these camera parameters remain unknown for all the methods.

In Figure 7d, the disparity computed by regular stereo matching algorithm is vulnerable to the ambiguity of defocus, especially in the region of carpet. The method from Li *et al.* [13] (Figure 7e) gives better result but with many wrong estimations since it can only handle equal apertures of both views. Our method (Figure 7f) delivers the best result and resolves the ambiguity much better. It is interesting to note that the ambiguity remains unsolved for regions appear in-focus in both views, i.e., the relative blur is close to zero in those regions.

Moreover, we evaluate our method on another real data set captured by the camera equipped in Nokia Lumia 1020 cell phone with f-number $f/2.2$. Images of both views are resized to 446×335. In Figure 8a and 8b, the left and right views focus at near and far sites, respectively. In Figure 8c, the disparity computed by regular stereo matching algorithm is vulnerable to the ambiguity of defocus. Our proposed method (Figure 8e) significantly improves the quality of disparity estimation over Li *et al.* [13] (Figure 8d) by taking advantage of the relation between relative blur and disparity.

5. Conclusions

We propose a blur-aware disparity estimation approach that can handle the mismatch of focus in stereo images. Unlike conventional depth from defocus techniques that model the relative blur within the focal stack, we exploit relative blur between stereo images to improve the disparity estimation. Our proposed method uses RBD model to characterize relative blur and disparity, which is more reliable than directly estimating the camera parameters. We have shown that the RBD model can be integrated into non-local cost aggregation framework for robust disparity estimation via compensating the data cost and aggregation weight. Experiments on both synthetic and real data confirm the effectiveness of our proposed approach.

References

[1] A. V. Bhavsar and A. N. Rajagopalan. Towards unrestrained depth inference with coherent occlusion filling. *International Journal of Computer Vision*, 97(2):167–190, Jul. 2011. 1

[2] M. Bleyer, C. Rhemann, and C. Rother. Patchmatch stereo - stereo matching with slanted support windows. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2011. 1

[3] F. Devernay, S. Pujades, and V. Ch.A.V. Focus mismatch detection in stereoscopic content. In *Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII, 82880E (February 6, 2012)*. 2, 5

[4] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions: empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, Aug. 2002. 4

[5] P. Favaro. Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 2

[6] P. Favaro, S. Soatto, M. Burger, and S. Osher. Shape from defocus via diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):518–531, 2008. 1

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second ed. Cambridge University Press, 2004. 3

[8] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2203–2141, 2011. 3

[9] E. Hecht. *Optics*. Fourth ed. Addison Wesley, 2002. 3

[10] R. T. Held, E. A. Cooper, and M. S. Banks. Blur and disparity are complementary cues to depth. *Current Biology*, 22:426–431, March 2012. 1

[11] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, Feb. 2013. 5

[12] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *International Journal of Computer Vision*, 63(3):175–189, Jul. 2005. 1

[13] F. Li, J. Sun, J. Wang, and J. Yu. Dual-focus stereo imaging. *Journal of Electronic Imaging*, 19(4), 2010. 1, 2, 5, 6, 7, 8

[14] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand. Defocus video matting. In *ACM SIGGRAPH*, 2005. 1

[15] C. Paramanand and A. N. Rajagopalan. Depth from motion and optical blur with an unscented kalman filter. *IEEE Transactions on Image Processing*, 21(5):2798–2811, May 2012. 1

[16] A. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:523–531, 1987. 3

[17] POV-Ray. Persistence of Vision Pty. Ltd., Williamstown, Victoria, Australia. http://www.povray.org/. 5

[18] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi. Depth estimation and image restoration using defocused stereo pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1521–1525, 2004. 1, 2

[19] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3):7–42, Apr. 2002. 1

[20] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 89(2):141–162, 2000. 1

[21] S. M. Seitz and S. Baker. Filter flow. In *IEEE International Conference on Computer Vision (ICCV)*, 2009. 1

[22] N. Shroff, A. Veeraraghavan, Y. Taguchi, O. Tuzel, A. Agrawal, and R. Chellappa. Variable focus video: Reconstructing depth and video for dynamic scenes. In *IEEE International Conference on Computational Photography (ICCP)*, 2012. 1

[23] W. Smith. *Modern Optical Engineering*. Third ed. McGraw-Hill, 2000. 3

[24] R. Szeliski and P. Golland. Stereo matching with transparency and matting. In *IEEE International Conference on Computer Vision (ICCV)*, 1998. 1

[25] Y. Takeda, S. Hiura, and K. Sato. Fusing depth from defocus and stereo with coded apertures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 1, 2

[26] Q. Yang. A non-local cost aggregation method for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2, 3, 4, 5, 6, 7, 8

[27] S. Zhu, L. Zhang, and H. Jin. A locally linear regression model for boundary preserving regularization in stereo matching. In *European Conference on Computer Vision (ECCV)*, 2012. 1