

Unconstrained Motion Deblurring for Dual-lens Cameras (Supplementary Material)

M. R. Mahesh Mohan, Sharath Girish, and A. N. Rajagopalan
 Indian Institute of Technology Madras
 {ee14d023, ee15b058, raju}@ee.iitm.ac.in

The contents of this supplementary material are arranged as follows:

1. Derivations and Proofs

- (a) Full derivation of DL pixel-mapping (in Eq. (4)).
- (b) Proof of Claim 2 (*i.e.*, the biconvexity property), and that individual optimizations in Eq. (8) are convex.
- (c) Derivation of equivalent LASSO formulation (in Eq. (9)).
- (d) Proofs of Remarks 1-3 (*i.e.*, the inadequacy of the normal camera methods [12, 18, 16] for DL-BMD).

2. Analysis and Discussions

- (a) Sensitivity of center-of-rotation (in narrow- and wide-FOV configurations).
- (b) Image-noise analysis.
- (c) Performance dependency with respect to depth quality.
- (d) Generalizability to diverse DL set-ups.
- (e) The uniqueness of our proposed DL pixel-mapping (Eq. (4)) over the homography mapping [16].
- (f) Implementation details and time analysis.

3. Additional Evaluations

- (a) Comparisons with all the methods [17, 9, 5, 1, 15, 11, 12, 18] for the examples in the main paper.
- (b) Additional examples (with extensive comparisons).
- (c) More real-world examples.

Note: The sections, equations and figures in the supplementary are indexed with a prefix ‘S’ (*e.g.*, Eq. (S11), Sec. S2).

S1. Derivations and Proofs

(a). Derivation of DL pixel-mapping: Using Eq. (1), the transformation of a world coordinate \mathbf{X} for a stationary camera (*i.e.*, $R = I$) can be written as

$$\mathbf{X}'' = \mathbf{X} + \mathbf{l}_b, \quad (\text{S1})$$

where $\mathbf{X}(3)$ ($= Z$) is the depth of the scene-point \mathbf{X} and \mathbf{l}_b is the baseline vector. The world coordinate \mathbf{X}'' maps to the corresponding sensor-coordinate \mathbf{x} (in accordance with Eq. (2)) as

$$\mathbf{x} = K^n \frac{\mathbf{X}''}{Z''} = K^n \frac{\mathbf{X}''}{Z} = K^n \frac{(\mathbf{X} + \mathbf{l}_b)}{Z} \quad \because \mathbf{l}_b(3) = 0. \quad (\text{S2})$$

Next, we consider the case of a camera pose-change R about the COR \mathbf{l}_c . That is, the world coordinate \mathbf{X} is transformed as

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b = R(\mathbf{X} + \mathbf{l}_b) + (I - R)\mathbf{l}_c + (I - R)\mathbf{l}_b. \quad (\text{S3})$$

Substituting Eq. (S2) in Eq. (S3) yields

$$\mathbf{X}' = ZR(K^n)^{-1}\mathbf{x} + (I - R)\mathbf{l}_c + (I - R)\mathbf{l}_b. \quad (\text{S4})$$

Using Eq. (2), the world coordinate \mathbf{X}' maps to the corresponding sensor-coordinate \mathbf{x}' as

$$\begin{aligned} \mathbf{x}' &= K^n \frac{\mathbf{X}'}{Z'} = \frac{Z}{Z'} K^n R(K^n)^{-1} \mathbf{x} + \frac{1}{Z'} K^n (I - R) \mathbf{l}_c + \frac{1}{Z'} K^n (I - R) \mathbf{l}_b, \\ &= \frac{Z}{Z'} \left(K^n R(K^n)^{-1} \mathbf{x} + \frac{1}{Z} K^n (I - R) \mathbf{l}_c + \frac{1}{Z} K^n (I - R) \mathbf{l}_b \right). \end{aligned} \quad (\text{S5})$$

As the sensor coordinate \mathbf{x}' is in homogeneous system, $\mathbf{x}'(3)$ should be unity. Therefore, the scale Z/Z' in Eq. (S5) can be considered as a normalization constant (say λ) that normalizes the third coordinate of \mathbf{x}' to 1, which establishes Eq. (4). ■

(b). Proofs of biconvexity:

(Note: For the proofs, we employ two well-known properties of convex functions: (1) Composite function of convex functions is convex. (2) Non-negative addition of convex (or biconvex) functions is a convex (or biconvex) function [3].)

Lemma S1: The costs L^n (and L^w) are individually biconvex in image \mathbf{I}_C^n and MDF \mathbf{w}^n (and \mathbf{I}_C^w and \mathbf{w}^w), respectively.

(Note: A function is biconvex in \mathbf{I} and \mathbf{w} if it is convex in \mathbf{I} for a given \mathbf{w} , and vice-versa.)

Proof: For a given \mathbf{I}_C^n , the cost L^n (in Eq. (5)) is given as

$$L^n = \|\mathbf{A}^n \mathbf{w}^n - \mathbf{I}_B^n\|_2^2 + \lambda_1^n \|\mathbf{w}^n\|_1 + \text{Constant}, \quad (\text{S6})$$

where the ‘Constant’ is $\lambda_2^n \|\nabla \mathbf{I}_C^n\|_1$. The first and second terms are composite of two convex functions (*i.e.*, a linear transformation of \mathbf{w}^n and its squared- l_2 or l_1 norm), and hence convex (Property 1). Further, the third term is convex as a constant is a convex function. Therefore, the cost L^n is convex with respect to \mathbf{w}^n (by Property 2 mentioned above).

For a given \mathbf{w}^n , the cost L^n (in Eq. (5)) can be equivalently represented as

$$L^n = \|\mathbf{M}^n \mathbf{I}_C^n - \mathbf{I}_B^n\|_2^2 + \lambda_2^n \|\nabla \mathbf{I}_C^n\|_1 + \text{Constant}, \quad (\text{S7})$$

where the ‘Constant’ is $\lambda_1^n \|\mathbf{w}^n\|_1$. Again, the first and second terms are composite of two convex functions (*i.e.*, a linear transformation of \mathbf{I}_C^n and its squared- l_2 or l_1 norm), and the third term is convex. Therefore, the cost L^n is convex with respect to \mathbf{I}_C^n (by Property 2). Hence L^n is biconvex in \mathbf{I}_C^n and \mathbf{w}^n . Similarly, we can show that L^w is biconvex in \mathbf{I}_C^w and \mathbf{w}^w . ■

Claim S1: The DL-BMD cost $L = L^n + L^w$ (Eq. (5)) is a *biconvex* function in image-pair $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ and MDF-pair $\{\mathbf{w}^n, \mathbf{w}^w\}$.

Proof: (We denote the function F for a given \mathbf{w} as $F_{\mathbf{w}}$.) From Lemma 1, the biconvexity of L^n implies

$$\begin{aligned} L_{\mathbf{w}^n}^n(\gamma \mathbf{I}_1^n + (1 - \gamma) \mathbf{I}_2^n) &\leq \gamma L_{\mathbf{w}^n}^n(\mathbf{I}_1^n) + (1 - \gamma) L_{\mathbf{w}^n}^n(\mathbf{I}_2^n) \quad \forall \{\mathbf{I}_1^n, \mathbf{I}_2^n, \mathbf{w}^n\}, \forall \gamma \in [0, 1]; \\ L_{\mathbf{I}_C^n}^n(\gamma \mathbf{w}_1^n + (1 - \gamma) \mathbf{w}_2^n) &\leq \gamma L_{\mathbf{I}_C^n}^n(\mathbf{w}_1^n) + (1 - \gamma) L_{\mathbf{I}_C^n}^n(\mathbf{w}_2^n) \quad \forall \{\mathbf{w}_1^n, \mathbf{w}_2^n, \mathbf{I}_C^n\}, \forall \gamma \in [0, 1]. \end{aligned} \quad (\text{S8})$$

As the cost L^n is independent of wide-angle parameters $\{\mathbf{I}_C^w, \mathbf{w}^w\}$, Eq. (S8) can be *equivalently* written as

$$\begin{aligned} L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n(\gamma \{\mathbf{I}_1^n, \mathbf{I}_1^w\} + (1 - \gamma) \{\mathbf{I}_2^n, \mathbf{I}_2^w\}) &\leq \gamma L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n(\{\mathbf{I}_1^n, \mathbf{I}_1^w\}) + (1 - \gamma) L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n(\{\mathbf{I}_2^n, \mathbf{I}_2^w\}); \\ L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n(\gamma \{\mathbf{w}_1^n, \mathbf{w}_1^w\} + (1 - \gamma) \{\mathbf{w}_2^n, \mathbf{w}_2^w\}) &\leq \gamma L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n(\{\mathbf{w}_1^n, \mathbf{w}_1^w\}) + (1 - \gamma) L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n(\{\mathbf{w}_2^n, \mathbf{w}_2^w\}). \end{aligned} \quad (\text{S9})$$

Eq. (S9) implies that L^n is biconvex in $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ and $\{\mathbf{w}^n, \mathbf{w}^w\}$. Following similar steps, the same inference can be derived for L^w too. Since the DL-BMD cost L is obtained by the summation of two biconvex function L^n and L^w , it must be biconvex with respect to image-pair $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ and MDF-pair $\{\mathbf{w}^n, \mathbf{w}^w\}$ (by Property 2). Hence proved. ■

Corollary S1: Introducing the DL prior in the DL-BMD objective L (Eq. (5)) retains the biconvexity property.

Proof: The prior $L^P = \alpha \|\mathbf{w}^n - \mathbf{w}^w\|_2 : \alpha > 0$ can be equivalently represented as

$$L^P = \alpha \|\mathbf{S} \mathbf{w}\|_2, \quad \text{where } \mathbf{S} = \text{diag}(I, -I), \text{ and } \mathbf{w} \text{ is the concatenation of vectors } \mathbf{w}^n \text{ and } \mathbf{w}^w. \quad (\text{S10})$$

First, $L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^P(\{\mathbf{I}_C^n, \mathbf{I}_C^w\})$ is a constant, and therefore it is convex with respect to dual image-pair $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$. Second, $L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^P(\{\mathbf{w}^n, \mathbf{w}^w\})$ is a composite of two convex functions, (*i.e.*, l_2 norm of a linear transformation of $\{\mathbf{w}^n, \mathbf{w}^w\}$).

Hence, the function is convex in $\{\mathbf{w}^n, \mathbf{w}^w\}$ (by Property 1). Therefore, the prior L^p is biconvex in $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ and MDF-pair $\{\mathbf{w}^n, \mathbf{w}^w\}$. Resultantly, the addition of L^p to a biconvex function L (Claim S1) retains the biconvexity property (by Property 2). ■

Corollary S2: Individual optimizations for MDF and latent image in Eq. (8) are convex.

Proof: MDF optimization problem is given as (Eq. (8)):

$$\tilde{\mathbf{w}}^n = \arg \min_{\mathbf{w}^n} \underbrace{\|\tilde{\mathbf{A}}^n \mathbf{w}^n - \mathbf{I}_B^n\|_2^2 + \alpha \|\mathbf{w}^n - \tilde{\mathbf{w}}^w\|_2^2}_{G} : \|\mathbf{w}^n\|_1 \leq \lambda_1^n, \quad (\text{S11})$$

The first term of G is convex in \mathbf{w}^n (proved in Lemma S1). As the DL prior (the second term) is convex with respect to *both* the MDFs $\{\mathbf{w}^n, \mathbf{w}^w\}$ combined, it should be convex in \mathbf{w}^n for a given \mathbf{w}^w . Also, the feasible set is convex [3]. Thus the MDF estimation is a convex optimization problem. The convexity of latent image estimation directly follows from the proof of Claim 1 (Eq. (S7)). ■

(c). Derivation of LASSO formulation for ego-motion estimation with prior: As Eq. (S11) is a convex optimization problem, $\hat{\mathbf{w}}^n$ is an optimal solution *iff*

$$\nabla G(\hat{\mathbf{w}}^n) = \mathbf{0} : \|\hat{\mathbf{w}}^n\|_1 \leq \lambda_3^n, \text{ where } \nabla G(\mathbf{w}^n) = 2 \cdot ((\tilde{\mathbf{A}}^{nT} \tilde{\mathbf{A}}^n + \alpha I) \mathbf{w}^n - (\tilde{\mathbf{A}}^{nT} \mathbf{I}_B^n + \alpha \tilde{\mathbf{w}}^w)) \quad (\text{S12})$$

Leveraging Eq. (S12), we frame a new optimization problem as follows: As ∇G is multi-dimensional, we consider the cost as the l_2 norm of ∇G (to convert to a single-valued objective function), *i.e.*,

$$\tilde{\mathbf{w}}^n = \arg \min_{\mathbf{w}^n} \|\nabla G(\mathbf{w}^n)\|_2^2 : \|\mathbf{w}^n\|_1 \leq \lambda_3^n \quad (\text{S13})$$

The new problem in Eq. (S13) possesses several desirable properties (in addition to the LASSO structure). (I) It is a convex optimization problem, *i.e.*, any local minima need to be a global minima and objective value of all minima should be the same. (II) $\|\nabla G(\mathbf{w}^n)\|_2^2 \geq 0$ and $\|\nabla G(\mathbf{w}^n)\|_2^2 = 0$ iff $\nabla G(\mathbf{w}^n) = \mathbf{0}$ (properties of norm). (III) Optimal solution $\hat{\mathbf{w}}^n$ of the problem in Eq. (S11) has to satisfy $\|\nabla G(\hat{\mathbf{w}}^n)\|_2^2 = 0$, which is a minima of Eq. (S13) (by Property (II)). Property (III) implies that all solutions of the optimization problem in Eq. (S11) will also be solutions of the LASSO framework. Also, since Eq. (S12) is a necessary and sufficient condition, all solutions of the LASSO framework will be solutions of the problem in Eq. (S11) (by Properties (I-II)), thereby establishing the equivalence of both the frameworks.

(d). Proofs of Remarks 1-3 in main paper:

(Note: For scenes with infinite depth, there exists *no* baseline-induced parallax (or depth cues). Thus, it is *finite* depths that are of primary interest to DL set-up as they induce *non-zero* disparities. Therefore, we consider scenes with finite-depth.)

Remark 1: The motion blur model of the methods [12, 18, 16] admits *only* a depth invariant model, whereas motion blur in a DL set-up warrants a depth variant model.

Proof: The pixel-mapping employed in single-lens model is given as [16]

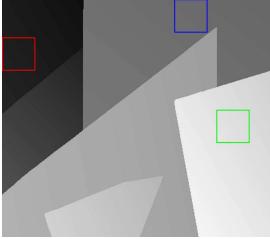
$$\hat{\mathbf{x}}' = \lambda K \hat{R} K^{-1} \mathbf{x} \quad (\text{S14})$$

where λ normalizes the third coordinate of $\hat{\mathbf{x}}'$ to 1. Note that the mapping in Eq. (S14) is *invariant* to scene-depth, unlike the depth-variant mapping of DL system due to baseline and COR (from Eq. (4)). Consider an arbitrary image-coordinate \mathbf{x}_0 with corresponding scene-depth Z_0 . Let $\hat{R}_{\{\mathbf{x}_0, Z_0\}}$ be the optimal pose in Eq. (S14) that *equates* to the homography-mapping of the DL system at \mathbf{x}_0 (*i.e.*, $\mathbf{x}' = \hat{\mathbf{x}}'$, for $\mathbf{x} = \mathbf{x}_0$). As Eq. (S14) is depth-invariant, fixing a pose suitable for \mathbf{x}_0 will *concurrently* fix the mapping at all other coordinates *irrespective* of their depth values. This clearly violates the inherent *depth-variant* mapping of DL system at those coordinates. Consequently, the PSF optimized for the coordinate \mathbf{x}_0 , through multiple poses $\hat{R}_{t\{\mathbf{x}_0, Z_0\}}$ for $t \in t_e$, *concurrently* fixes the PSF at other coordinates *irrespective* of their depth values; thereby failing to model the depth-variant PSFs in a DL setup. ■

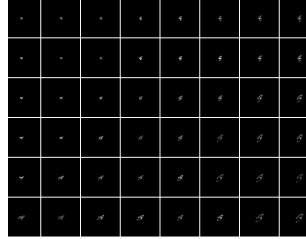
Remark 2: The blur model of the methods [12, 18, 16] modulate the baseline with camera poses, but it must be independent for a DL set-up (for scene-consistent disparities).

Proof: Motion blur model in single-lens system is given as [16]

$$\hat{\mathbf{I}}_B^n = \frac{1}{t_e^n} \int_{t_e^n} P^n(\hat{R}_t(\mathbf{Y})) dt, \quad (\text{S15})$$



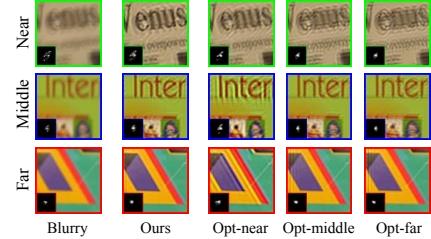
(a) Scene depth



(b) Blur kernel (PSF)

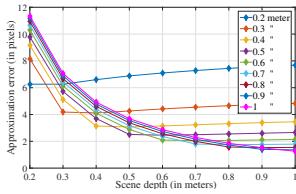


(c) Blurred image

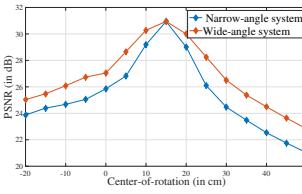


(d) Image-patches

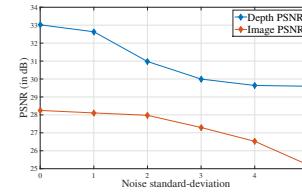
Figure S1. DL configuration warrants a *depth-variant* transformation (Note the variation of PSF in Fig. (b) with respect to the scene depth in Fig. (a)). As the single-lens motion blur model is *depth-invariant*, the model optimized for a fixed depth can fail for other depths, leading to *ineffective* deblurring across depths (Fig. (d)).



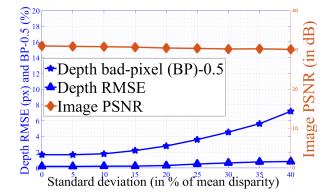
(a) Model inaccuracies



(b) Sensitivity-analysis of COR



(c) Image-noise analysis



(d) Depth-noise analysis

Figure S2. Analysis: (a) Model inaccuracies of the homography model. (b) Sensitivity of COR: Both narrow-angle and wide-angle configurations are very sensitive to COR, with the former exhibiting relatively more sensitivity. (c-d) Effect of image and depth noise.

where the world-coordinate system \mathbf{Y} is defined with respect to the optical center (*i.e.*, $\mathbf{l}_b = \mathbf{l}_c = 0$). In the DL blur model (Eqs. (2)-(3)), the effect of stereo baseline is *independent* of camera pose-changes, and the disparity relation between stereo image-pair is due to the baseline (\mathbf{l}_b). Enforcing the single-lens model in the narrow-angle image leads to $\hat{\mathbf{I}}_B^n = (1/t_e^n) \int_{t_e^n} P^n(\hat{R}_t(\mathbf{X} + \mathbf{l}_b)) dt$ (where the world coordinate \mathbf{X} is defined with respect to wide-angle camera, as followed in Eqs. (1)-(2)). Evidently, the effect of baseline in this case *varies* with pose-change \hat{R}_t , *unlike* the DL model. Specifically, it characterizes an alien dual-lens set-up with its own physical lens-separation, and in turn the scene disparities, getting modulated by pose-changes over time. ■

Remark 3: The methods [12, 18, 16] also admit the ill-posedness that disrupts scene-consistent disparities.

This is a special case of Claim 1 (in the main paper), wherein $\mathbf{l}_b = \mathbf{l}_c = 0$.

S2. Analysis and Discussions

(a). Sensitivity of center-of-rotation: To analyze the sensitivity of COR for narrow-angle and wide-angle configurations, we considered images blurred with a common COR, and performed deblurring by perturbing the COR vector and using the true ego-motion (*identically* for both the configurations). Figure S2(b) compares the average PSNR of deblurred images for different COR approximations. The figure clearly shows a significant drop in deblurring performance as the approximated COR deviates from the true COR. Also, note the detrimental effect of the common COR approximation about the camera center (that is followed in single-lens BMD methods). The figure also reveals higher sensitivity of COR in narrow-angle configuration as shown by the higher rate of its performance-drop. This is due to higher focal-length, and hence larger blur inherent in narrow-angle setup which is a function of COR (as noted in Sec. 2).

(b). Image-noise analysis: To analyze the effect of noise in our DL-BMD method, we experimented with blurry images corrupted with additive white Gaussian noise. Standard-deviation of noise (in pixels) is varied from 0 (noise-less case) to 5. Fig. S2(c) plots the average PSNRs of a deblurred image and depth estimate corresponding to different noise levels. The average PSNRs for deblurred image and depth-estimate is more than 25 dB and 29 dB, respectively, over the entire standard-deviation range; this clearly reveals the noise-robustness of our algorithm. Although we did *not* perform denoising in any examples, for *very* high noisy levels, the blurred image-pair need to be denoised prior to deblurring. This is because noise can deteriorate image-gradients which are required for ego-motion estimation (Sec. 4).

(c). Performance dependency with respect to depth quality: The total variation prior in the DL-BMD cost is employed to curb ringing artifacts. We analysed the depth-dependency by adding AWGN in disparity-map (following [10, 8, 13]). The

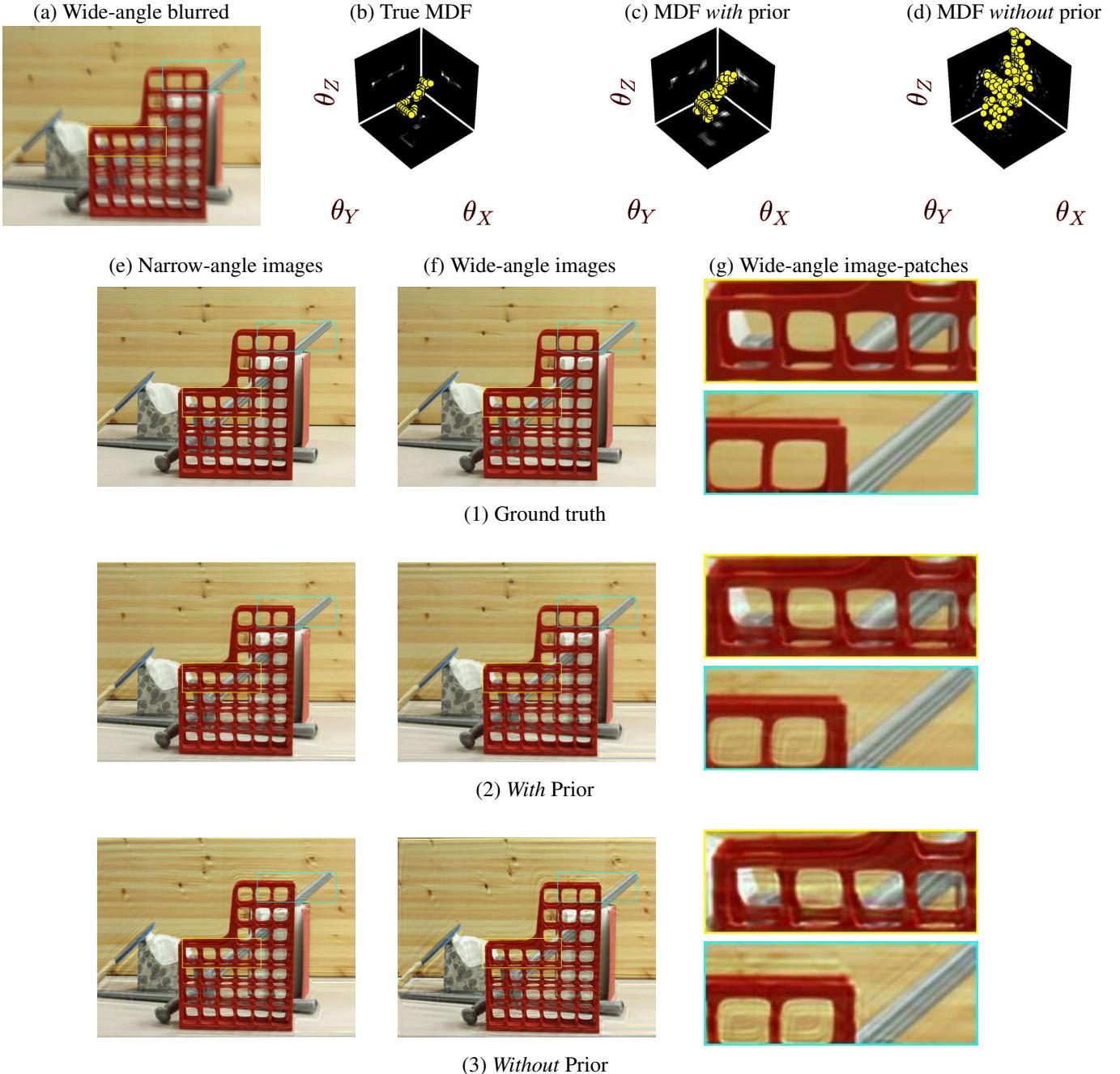


Figure S3. Effect of the proposed prior: The MDF estimate and the deblurred image-patches of prior-less case clearly show a significant rotational ambiguity (Figs. [g,3] and d). Also, the deblurred image in the prior-less case exhibits considerable ringing artifacts and residual blur (Figs. [e-f,3]), which could be possibly due to the less accurate MDF estimate (Fig. d). In contrast, the addition of the prior successfully curbs the pose ambiguity (Figs. [g,2] and c), improves the MDF accuracy (Fig. c) and produces better deblurring quality (Figs. [e-f,2]).

standard deviation (SD) of noise is varied from 0 to 40% of mean disparity, in the disparity-map estimate in *all* iterations (a worst-case scenario). In Fig. S2(d), we plot the results averaged over five trials for each SD-unit (for the example in Fig. S3), where we utilize the metrics RMSE and bad pixel ratio for depth. Note that over the entire SD-range, image-PSNR and depth-RMSE are reduced by only 0.875 dB and 0.622 px, respectively, which clearly reveals our method's robustness.

(d). Generalizability to diverse DL set-ups: Here, we show the generalizability of our algorithm to different types of DL

Configuration	Metrics	Blurred	Xu <i>et al.</i> [17]	Mohan et al. [9]	Ours
Narrow-Narrow	Image PSNR/IFC/VIF	27.27 / 1.75 / 0.23	19.90 / 1.08 / 0.22	29.21 / 2.30 / 0.36	31.03 / 3.04 / 0.43
	Depth PSNR	29.22	15.83	29.50	30.35
Narrow-Wide	Image PSNR/IFC/VIF	27.33 / 1.78 / 0.23	19.86 / 1.13 / 0.22	26.50 / 1.95 / 0.31	30.50 / 3.10 / 0.42
	Depth PSNR	28.51	15.29	28.56	31.11
Wide-Wide	Image PSNR/IFC/VIF	27.87 / 1.97 / 0.27	14.56 / 0.94 / 0.17	25.90 / 2.04 / 0.32	30.64 / 4.40 / 0.56
	Depth PSNR	30.15	13.88	28.56	30.62

Table S1. Generalizability to diverse DL set-ups: Our method consistently outperforms the methods of [17, 9] in the PSNR, IFC and VIF metrics for image and the PSNR metric for depth.

set-up. The image PSNR, VIF, IFC metrics and depth PSNR metric are shown in the Table S1 for the three DL-configurations: Narrow-Narrow, Narrow-Wide, Wide-Wide. We consider the same exposure time for both cameras, 52mm focal length for narrow angle camera and 26mm focal length for the wide angle camera. The values reported in the Table S1 are averaged over three examples. As can be seen, our method performs consistently better than the methods of [17, 9] in all three configurations. Specifically, in terms of Image/Depth PSNR, our method outperforms [9] by 0.82/0.85 dB for Narrow-Narrow setup, 4.00/2.55 dB for Narrow-Wide setup and 4.74/2.06 dB for Wide-Wide setup.

(e). **Uniqueness of our proposed DL pixel-mapping (Eq. (4)) over homography mapping [16].** The uniqueness is due to the latter's *depth-invariant* nature, which we illustrate with two experiments. For a camera-pose sampled from a real trajectory [6], Fig. S2(a) shows the *best* approximation error of the homography-mapping of normal cameras to the pixel-mapping of Eq. (4), optimized for a given depth (under least-squares criteria), for the set-up of Samsung S9+. Notably, the homography *fails* to model the DL pixel-wise mapping *consistently over different depths*, which clearly illustrates the uniqueness of Eq. (4). Further, to analyze the depth-variant nature of PSFs, Figs. S1(a-d) consider a camera trajectory and a 3D scene from [14]. Fig. S1(b) shows the corresponding PSFs (projected using Eq. (4)), which reveals depth-dependency of blur, with lower depths exhibiting severe blurs relative to the farther ones. Figure S1(d) shows the deblurred image-patches for different depths employing the normal camera method [18], optimized for a given depth; it is evident that this approach is *not* quite successful due to the *depth-dependency* of the blur, which clearly necessitates a new approach for DL-BMD.

(f). **Implementation details and time analysis:** We used a PC with an Intel Xeon processor and a 16 GB RAM for all experiments, and implemented our algorithm in MATLAB. For the scale-space based alternating minimization, we used 5 scales with 6 iterations each. The scaling factor for the i th scale is selected as $\frac{1}{\sqrt{2}}^{(i-1)}$. For estimating COR (following Eq. (7)), we have employed the MATLAB built-in function `lsqnonlin`. For depth estimation, we adopted the optical-flow algorithm of [7] and we employed the default parameters (as it provides a good trade-off between speed and accuracy). For optimizing the cost for MDF (Eq. (9)), we used the LARS solver of [4] (which efficiently solves LASSO problems). The regularization for the proposed MDF-prior α is adapted with the scales as $5^{\frac{(9-i)}{2}}$ (Note that a higher regularization is employed as MDF vectors have smaller values as compared to images). We have selected the sparsity regularization (λ_3) as 0.01 for both narrow-angle and wide-angle MDFs. We employed ADMM [2] to optimize the cost for latent-image with total-variation prior, where we used the total-variation regularization as 0.005. For latent image estimation, we consider grey-scale image until the final scale and 5th iteration (to reduce the computational time). We found that for deblurring a 1280×720 RGB narrow-angle image (of maximum blur-length of 30 pixels) and a focal-length ratio of two, our unoptimized MATLAB implementation took about 23 minutes to deblur the dual image-pair. A detailed break-up of the time taken for the final scale, final iteration is as follows: optimizing COR took 49.7s, estimating depth took 14.7s, MDF estimation took 56.4s, and RGB latent image estimation took 39.4s.

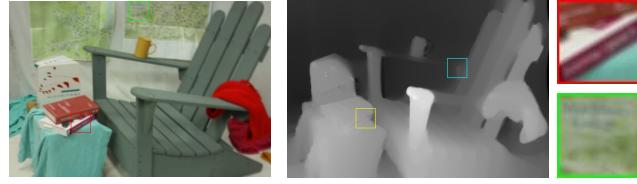
S3. Additional Evaluations

We employed the publicly available codes for [17, 18, 15, 11, 12, 5]. Codes for [1] and [9] were shared by the authors upon request. We used default parameters for all the codes. Figure S3 provides the full images corresponding to the Fig. 1. Figures S4-S7 provide a thorough qualitative evaluation with all the competing methods, *i.e.*, state-of-the-art DL-BMD [17], single-lens BMD [18, 12], depth-aware BMD [5, 1], light field BMD [9], and deep learning based approaches [15, 11]. In addition to the examples presented in the main paper, we provide eight examples which include five real scenes under low-light and well-lit scenarios (Figs. S11, S12, and S13). For each example, we show deblurred narrow-angle image (with highlighted patches magnified), and estimated depth (with highlighted patches). The consistent performance of our algorithm

over all the examples in retrieving scene-features at different depths (unlike the competing methods), with negligible residual blur and ringing artifacts, and accurate depth estimates amply demonstrates the superiority of our method.

References

- [1] M. Arun, A. N. Rajagopalan, and Gunasekaran Seetharaman. Multi-shot deblurring for 3d scenes. In *IEEE International Conference on Computer Vision (ICCV) Workshops*, pages 19–27, 2015. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. [6](#)
- [3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. [2](#), [3](#)
- [4] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004. [6](#)
- [5] Zhe Hu, Li Xu, and Ming-Hsuan Yang. Joint depth estimation and camera shake removal from single blurry image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2893–2900, 2014. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [6] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European Conference on Computer Vision (ECCV)*, pages 27–40. Springer, 2012. [6](#)
- [7] Ce Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. [6](#)
- [8] Lee-Kang Liu, Stanley H Chan, and Truong Q Nguyen. Depth reconstruction from sparse samples: Representation, algorithm, and sampling. *IEEE Transactions on Image Processing*, 24(6):1983–1996, 2015. [4](#)
- [9] M. R. Mahesh Mohan and A. N. Rajagopalan. Divide and conquer for full-resolution light field deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6421–6429, 2018. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [10] Srimanta Mandal, Arnav Bhavsar, and Anil Kumar Sao. Depth map restoration from undersampled data. *IEEE Transactions on Image Processing*, 26(1):119–134, 2016. [4](#)
- [11] T. M. Nimisha, Akash Kumar Singh, and A. N. Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *IEEE International Conference on Computer Vision (ICCV)*, pages 4762–4770, 2017. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [12] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1628–1636, 2016. [1](#), [3](#), [4](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [13] Gernot Riegler, Matthias Rüther, and Horst Bischof. Atgv-net: Accurate depth super-resolution. In *European conference on computer vision*, pages 268–284. Springer, 2016. [4](#)
- [14] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. [6](#)
- [15] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8174–8182, 2018. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [16] Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision (IJCV)*, 98(2):168–186, 2012. [1](#), [3](#), [4](#), [6](#)
- [17] Li Xu and Jiaya Jia. Depth-aware motion deblurring. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2012. [1](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)
- [18] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1107–1114, 2013. [1](#), [3](#), [4](#), [6](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#)



(a) Blurry image



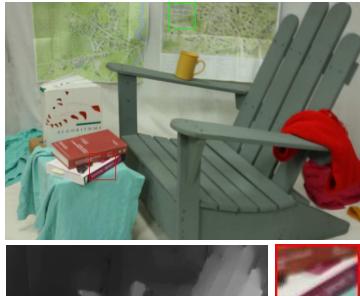
(b) Ours



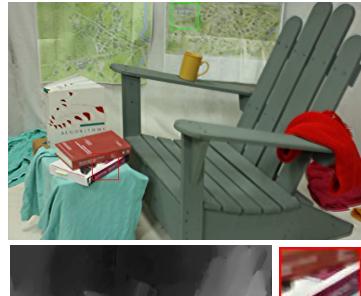
(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



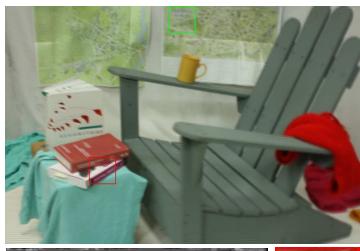
(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]



(f) Hu *et al.* [5]

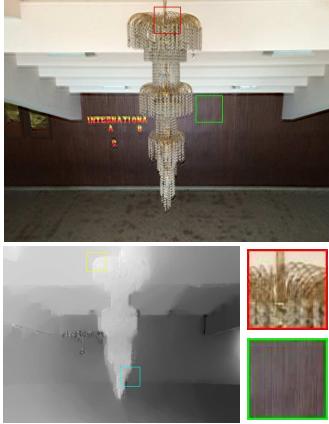


(g) Nimisha *et al.* [11]

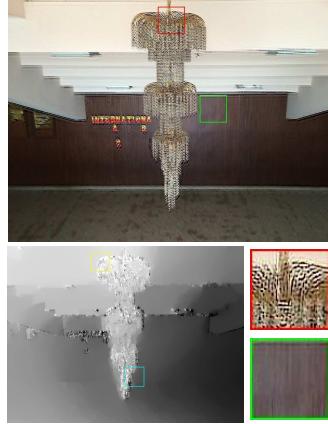
Figure S4. Synthetic experiments: Our method is able to retrieve the finer details at different depth levels with little ringing. The text in the patches are sharper when compared with the other methods.



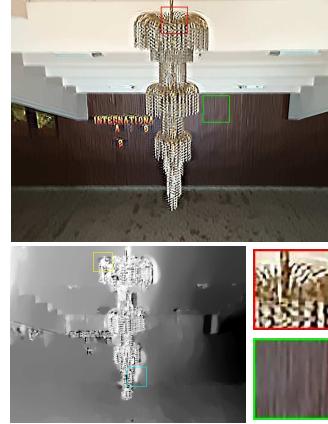
(a) Blurry image



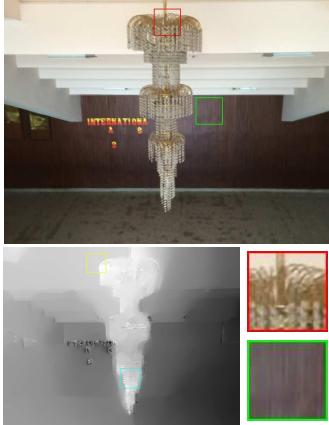
(b) Ours



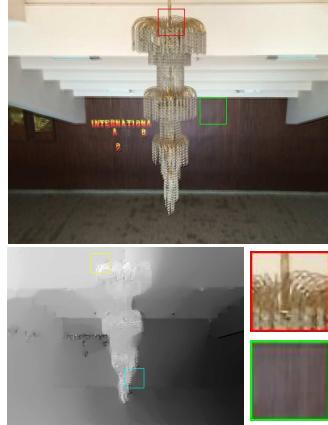
(c) Mohan *et al.* [9]



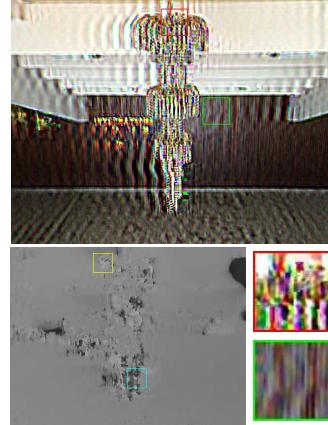
(d) Pan *et al.* [12]



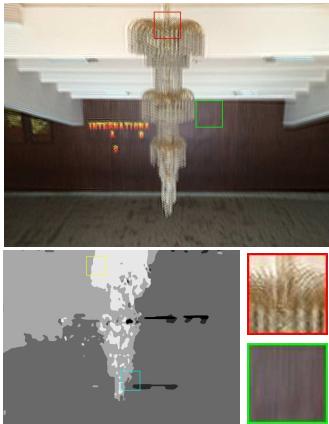
(e) Tao *et al.* [15]



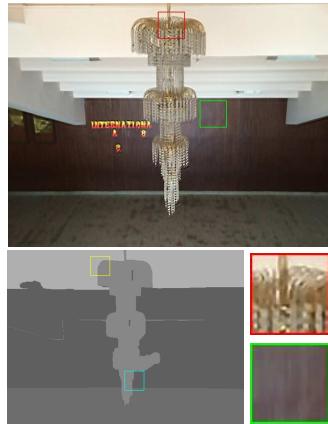
(f) Xu *et al.* [18]



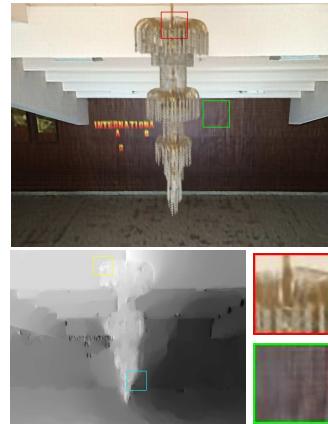
(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]



(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S5. Real experiments: Unlike the results of [9, 12, 17], our method is able to deblur the image with no artifacts while also recovering the finer details on the chandelier and the wall. Also, note the ineffectiveness of the single-lens methods [5, 12, 18] in DL configuration.



(a) Blurry image



(b) Ours



(c) Mohan *et al.* [9]



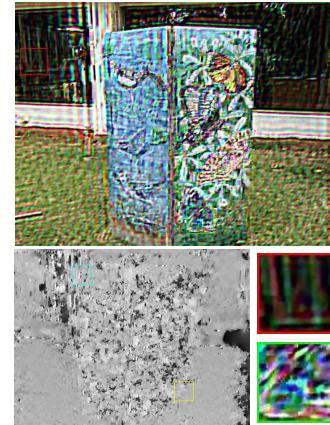
(d) Pan *et al.* [12]



(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]



(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S6. Real experiments: The results of the deep learning methods of [11, 15] exhibit a significant amount of residual blur. Our method is able to faithfully preserve the depth information and exhibits the least ringing artifacts.



(a) Blurry image



(b) Ours



(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



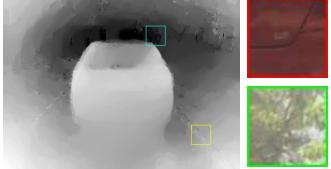
(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



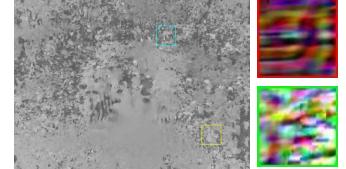
(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]

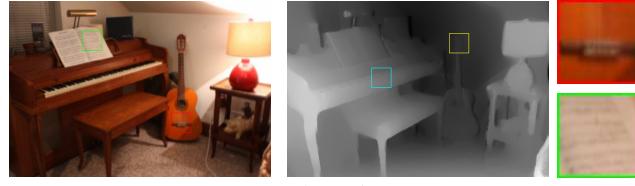


(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

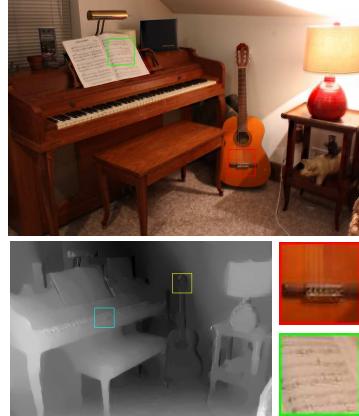
Figure S7. Real experiments: Unlike all the other methods, our method is able to recover fine textual information on the car and also the thin branches in the background. The light field method of [9] and the dual lens method of [17] exhibit severe artifacts in the deblurred image.



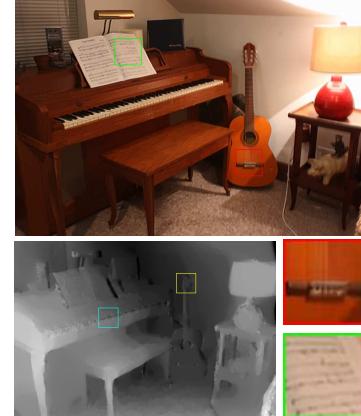
(a) Blurry image



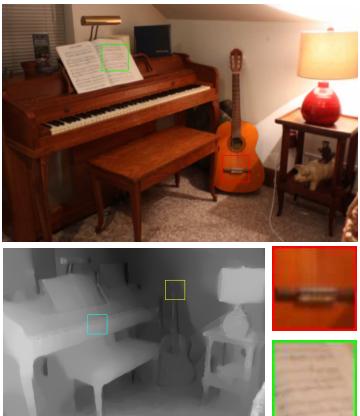
(b) Ours



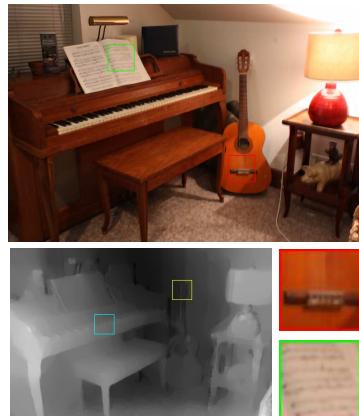
(c) Mohan *et al.* [9]



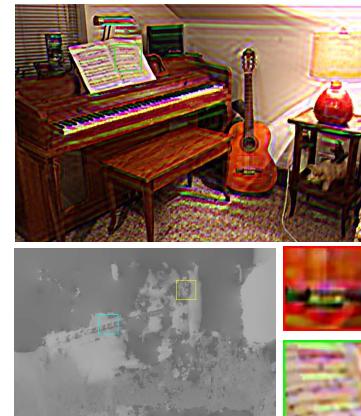
(d) Pan *et al.* [12]



(e) Tao *et al.* [15]



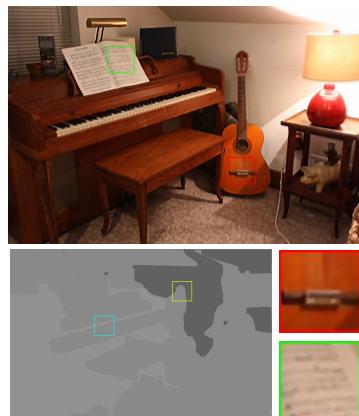
(f) Xu *et al.* [18]



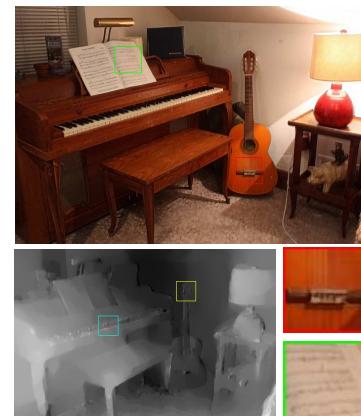
(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]

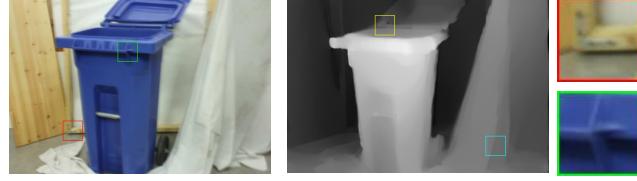


(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S8. Synthetic experiments: Our method recovers the sharp details of the image such as on the guitar and the book without any erroneous depth values.



(a) Blurry image



(b) Ours



(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]



(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S9. Synthetic experiments: Our method retrieves scene features without introducing artificial structures (unlike the deep learning method [11]), e.g., the features in the highlighted patches in [11] are hallucinated by the deep learning N/W. Resultantly, depth estimates of [11] are inadequate due to uneven hallucinations in the dual image-pair.



(a) Blurry image



(b) Ours



(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]

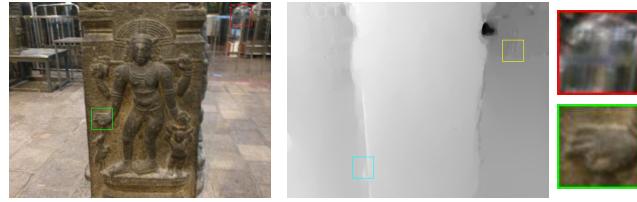


(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S10. Synthetic experiments: Our method has considerably lesser ringing and sharper features (such as on the closer doll and the farther basket) leading to good depth estimates compared to the other methods.



(a) Blurry image



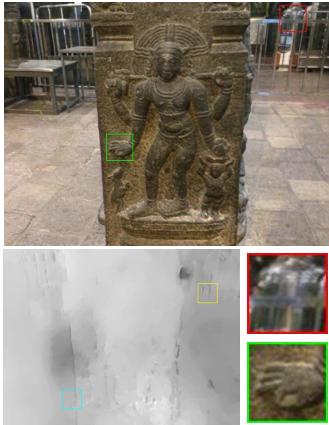
(b) Ours



(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



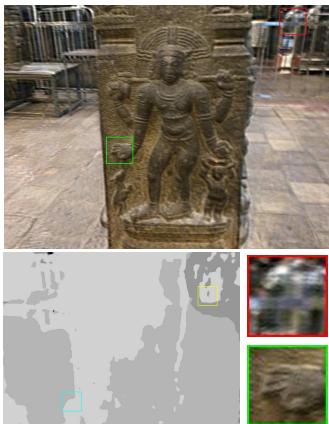
(e) Tao *et al.* [15]



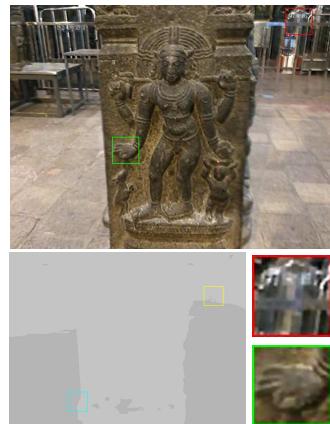
(f) Xu *et al.* [18]



(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]

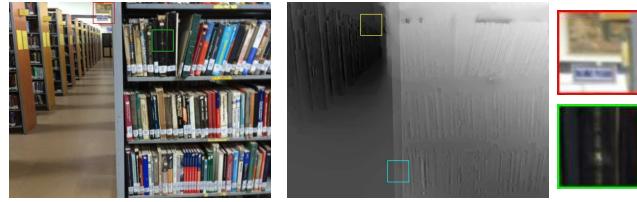


(f) Hu *et al.* [5]

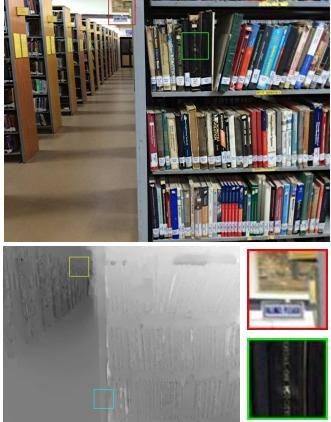


(g) Nimisha *et al.* [11]

Figure S11. Real experiments: Even in a low-light (noisy) scenario, the uniform deblurring performance of our method over different depth levels reveals the noise-robustness of our method.



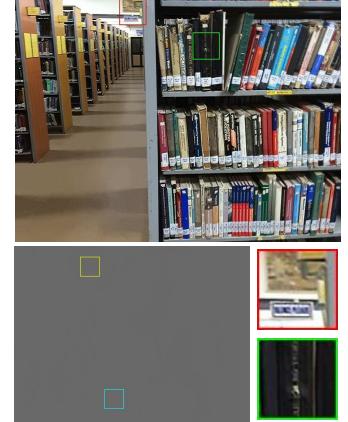
(a) Blurry image



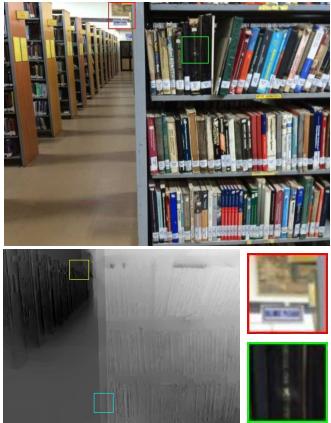
(b) Ours



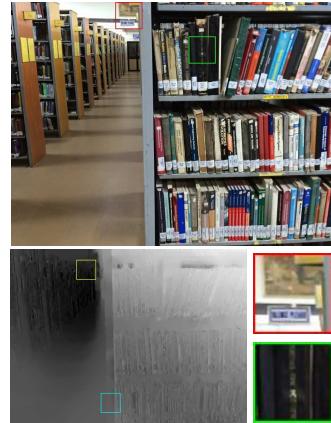
(c) Mohan *et al.* [9]



(d) Pan *et al.* [12]



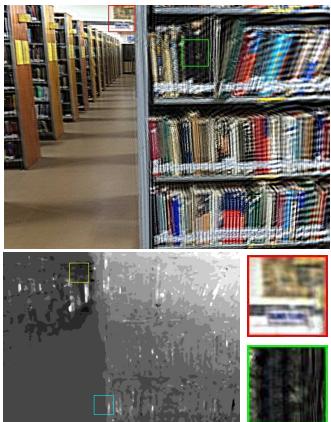
(e) Tao *et al.* [15]



(f) Xu *et al.* [18]



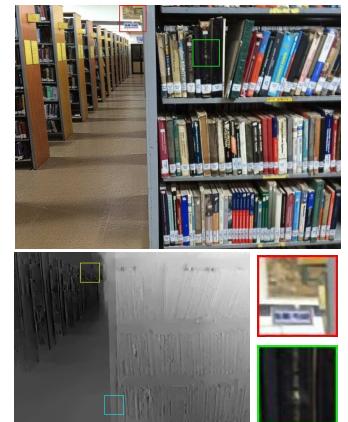
(g) Xu *et al.* [17]



(e) Mathamkode *et al.* [1]



(f) Hu *et al.* [5]



(g) Nimisha *et al.* [11]

Figure S12. Real experiments (Well-lit scenario): The uniform deblurring over different depth levels yet again proves the effectiveness of our proposed method. Notably, the depth estimate is more accurate and finer in our approach as compared to the competing methods.

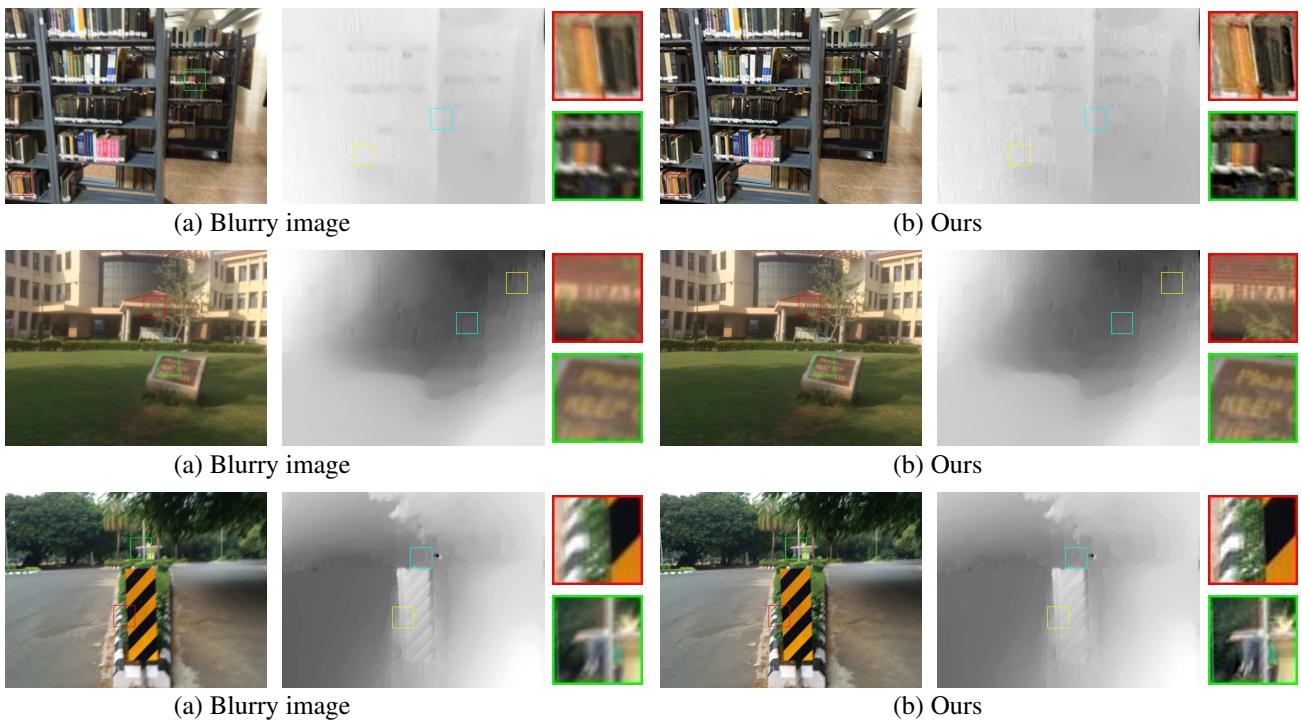


Figure S13. More real-world examples: (a) Blurry Image and (b) Our result.