Unconstrained Motion Deblurring for Dual-lens Cameras

M. R. Mahesh Mohan, Sharath Girish, and A. N. Rajagopalan

Indian Institute of Technology Madras

{ee14d023,ee15b058,raju}@ee.iitm.ac.in

Abstract

Recently, there has been a renewed interest in leveraging multiple cameras, but under unconstrained settings. They have been quite successfully deployed in smartphones, which have become de facto choice for many photographic applications. However, akin to normal cameras, the functionality of multi-camera systems can be marred by motion blur which is a ubiquitous phenomenon in hand-held cameras. Despite the far-reaching potential of unconstrained camera arrays, there is not a single deblurring method for such systems. In this paper, we propose a generalized blur model that elegantly explains the intrinsically coupled image formation model for dual-lens set-up, which are by far most predominant in smartphones. While image aesthetics is the main objective in normal camera deblurring, any method conceived for our problem is additionally tasked with ascertaining consistent scene-depth in the deblurred images. We reveal an intriguing challenge that stems from an inherent ambiguity unique to this problem which naturally disrupts this coherence. We address this issue by devising a judicious prior, and based on our model and prior propose a practical blind deblurring method for dual-lens cameras, that achieves state-of-the-art performance.

1. Introduction

Motion blur due to camera shake is a ubiquitous phenomenon in hand-held photography. The challenging problem of blind motion deblurring (BMD) deals with estimating a clean image from a *single* motion blurred observation. Since most computer vision works are designed for blur-free images, BMD is a continuing research endeavour, replete with several theories and methods [5, 23, 40, 26, 48].

Most modern cameras come with dual-lens (DL) configuration, that can have *different or identical* focal lengths or field-of-views (FOVs), exposure times, and image resolutions (which we refer to as unconstrained set-up). Works already exist that have generously invoked such flexibility. For example, HDR imaging [29, 2, 43], low-light photography [47], and stereoscopics [30] require differently-exposed

stereo images with overlapping exposure times; whereas in super-resolution [14] and visual odometry [24, 13] the stereo images are captured with nearly-identical exposure times. The world of smartphones is today experiencing a proliferation of unconstrained DL cameras, wherein almost all devices consider a narrow-FOV camera paired to a conventional wide-FOV camera (for portrait photography), with possibly different resolutions. Akin to normal cameras ([12, 56, 22, 32]), images captured with DL cameras are also susceptible to motion blur. However, there does *not* exist a single BMD method that addresses the growing trend of unconstrained DL set-up.

The problem of BMD for DL cameras is fraught with additional challenges over those present in normal cameras. First, a DL set-up warrants deblurring based on scene depth [51], whereas methods for normal cameras are typically independent of depth [26, 52, 48, 9], as recovering scene depth from a single blurred image is a difficult problem [12, 9]. Second, any method for DL-BMD must ensure scene-consistent disparities in the deblurred image-pair (akin to angular coherence in light fields [23, 40]), which also incidentally opens up many potential applications [14, 29, 37, 24]. This is an additional conformity condition in DL-BMD. The narrow-FOV genre popularized by current smartphones admits further issues. The higher focal length of narrow-FOV camera amplifies the effect of camera shake [48], and thereby renders motion blur *more* severe. Moreover, the assumption of center-of-rotation (COR) of the camera at the optical center significantly affects ego-motion estimation, and hence the deblurring quality [12, 10]. In practice, COR may be located at a point far away, such as in the photographer's wrist in case of handheld shake [39, 15]. It must be noted that *none* of the existing BMD methods are designed to handle the COR issue. The higher focal length exacerbates the issue of COR as well in DL set-up.

Traditional BMD methods for normal cameras restrict themselves to space-invariant blur [6, 53, 55, 46, 57, 41]. Whyte *et al*. [48] showed that motion blur in general is *space-variant* and is primarily caused by camera rotations. This is predominantly followed in later methods

[26, 42, 52]. However, it is shown in [18, 5, 23, 40, 51] that the BMD methods developed for normal cameras are *seldom* successful for computational cameras. This has necessitated new methods that adhere to the modified camera-principles and ensure coherencies in the computational data [23, 40, 51]. For the case of DL cameras, Xu *et al.* [51] restrict to a constrained set-up, *i.e.*, require two identical cameras to work in synchronization, so that the *same* blur applies to both images. Importantly, the method imposes strong assumptions on blur that it is *primarily* caused by inplane translations (which does *not* hold good in practice [48]), and that scene is fronto-parallel with layered depth. Recently, DL video deblurring methods have been proposed [28, 34], but they address dynamic objects and necessitate as input *multiple* stereo image-pairs.

For the case of light field cameras, existing methods constrain all multi-view images to share identical camera settings and ego-motions [18, 5, 23, 40]. Though this property is inherent to light field cameras due to the micro-lens set-up, it need *not* hold for an unconstrained set-up. Also, the imaging principle of light field is quite different due to the lens effect [5, 23]. Importantly, none of the methods (except [23]) is applicable to our problem because their objective function warrants 4D light field [40] or multi-view images with *identical* FOVs, resolutions, and exposure times for latent image [18] (or texture [5]) update.

Among other closely related works, Hu *et al.* [11] estimate a clean image and layered depth from a *single* blurred image. However, [11] requires the blur to be primarily due to inplane translations. To reduce the ill-posedness, Pan *et al.* [27] assume that *accurate* depth is known a priori, but this is difficult to achieve in blur scenarios [18, 12]. Further, the method imposes strong assumption of *uniform* ego-motion parameterized by a *single* camera-pose that has *negligible* rotation, which is very unlikely in practice [16, 42, 48]. Mathamkode *et al.* [1] propose a method for multi-shot BMD, but employ four images and restrict to layered depth scenes. Moreover, [1] requires all the images to be registered within a few pixels (which is possible in ego-motion induced disparities [41], but does *not* hold good for baseline induced disparities [3]).

In this paper, we address the hitherto unaddressed problem of BMD for *unconstrained* DL set-up. First, we propose a *DL-blur model* that accounts for arbitrary camera settings and COR. Second, we reveal an *inherent ill-posedness* present in DL-BMD, under the unconstrained exposure scenario ([47, 29, 30, 43, 49, 54]), that disrupts scene-consistent disparities. To this end, we devise a *new prior* that respects consistency of disparities (and also aids ego-motion estimation). Priors that render the resultant cost highly nonconvex or warrant a costly optimization are *not* desirable [40, 26, 52]. We show that our prior is convex and retains the *biconvexity* property (required for convergence

[31, 52, 6]) and is amenable to the *efficient* LASSO framework. Finally, based on the proposed model and prior, we develop a practical DL-BMD method. It eliminates the restrictions of [23, 11, 51] and also addresses the COR issue. To mitigate the processing difficulties incurred in jointly optimizing multiple images or ego-motion, we propose a strategy that decomposes the high-dimensional BMD problem into subproblems, while enforcing the prior and convexity. Our main contributions are summarized below:

- This is the first attempt to formally address blind motion deblurring for unconstrained camera configurations. To this end, we introduce a *generalized* DL blur model, that also allows for arbitrary COR.

- We reveal an inherent *ill-posedness* present in DL-BMD, that disrupts scene-consistent disparities. To address this, we devise a prior that ensures the biconvexity property and admits efficient optimization.

- Employing the introduced model and prior, we propose a practical DL-BMD method that achieves state-of-the art performance for current DL set-up. It ensures scene-consistent disparities, and accounts for the COR issue (which is a first for BMD).

2. Motion Blur model for Unconstrained DL

In this section, we introduce a DL motion blur model and its corresponding pixel-wise mapping, considering cameras with different FOVs, exposure times, and resolutions.

In a DL camera set-up, at any instant of time, one camera will perceive a shifted world (by the stereo baseline) with respect to that of a reference camera. Following [23, 26, 42, 52, 48], we consider a blurred image as the integration of rotation-induced projections of world over the exposure time, the rotations being caused by camera shake, but do *not* constrain the COR to be *only* at the optical center. Thus, a rotational pose-change translates a world coordinate $\mathbf{X}$ to

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}, \tag{1}$$

where R is the corresponding rotational matrix [48], $\mathbf{l_b}$ is the baseline vector ($\mathbf{l_b} = \mathbf{0}$ for the reference camera) and $\mathbf{l_c}$ is the unconstrained COR vector (defined in the world coordinate system). We indicate the parameters of the relatively narrow-angle camera by superscript n and the other by superscript w. Thus a DL motion blurred image-pair ($\mathbf{I}_B^w$ and $\mathbf{I}_B^n$) (with the COR factored in) can be represented as

$$\mathbf{I}_B^w = \frac{1}{t_e^w} \int_{t \in t_e^w} P^w\big(R_t(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c}\big) \, dt,$$
$$\mathbf{I}_B^n = \frac{1}{t_e^n} \int_{t \in t_e^n} P^n\big(R_t(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}\big) \, dt, \tag{2}$$

where the wide-angle camera is considered as reference (without loss of generality). In practice, the COR (l_c) remains fixed over the exposure time (t_e) [12].

For sake of simplicity, with a slight abuse of notation, we use $P^n(\cdot)$ and $P^w(\cdot)$ to denote DL images formed by projecting the world onto the narrow- and wide-angle camera sensors, respectively, that is, by the argument of $P\big(R_t(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}\big)$ we mean a transformation mapping $T_{(R_t, \mathbf{l_c}, \mathbf{l_b})} : \mathbf{X} \to R_t(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}$, $\forall \mathbf{X}$ in world-space. In general, a given world coordinate $\mathbf{X_0}$ is mapped to a (homogeneous) sensor coordinate $\mathbf{x_0}$ in accordance with $\mathbf{x_0} = K\mathbf{X_0}/Z_0$, where Z_0 is the scene depth and K is the intrinsic camera matrix ($K = \text{diag}(f, f, 1)$, and f is the focal length in pixels). Note that different image resolutions are captured by the scale factors that are used to convert parameters from metres to pixels [48]. Resultantly, for a world coordinate $\mathbf{X_0}$, it is evident from Eq. (2) that the pixel-displacement due to camera motion (or $R_t \mathbf{X_0}$) and COR (or $\mathbf{l_c} - R_t \mathbf{l_c}$) gets relatively amplified in narrow-angle camera by a factor of f^n/f^w. (Typical values of f^n/f^w are around two in portrait-enabled smartphones, and hence exacerbates the issues of motion blur and COR).

To linearize the dual-lens motion blur model, we *equivalently* represent Eq. (2) as the integration of image-projections over pose-space (instead of over time) as

$$\mathbf{I}_B^n = \int_{p \in \mathbb{P}^3} w^n(p) \cdot P^n\big(R_p(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}\big)\, dp, \quad (3)$$

where $\mathbb{P}^3$ is the 3D space covering plausible rotational camera poses. The quantity $w^n(p_0)$ gives the fraction of exposure time over which the camera stayed in pose p_0, which defined over the entire $\mathbb{P}^3$ is referred to as motion density function (MDF). The MDF formulation can accommodate both regular and irregular camera motion (unlike [18, 40, 42]). The consideration of full 3D rotations accommodates *both* narrow- and wide-FOV cameras [42].

We now proceed to derive the pixel-mapping in DL set-ups. This is the counterpart of homography-mapping in normal cameras (as discussed in [48]), which is extensively used to create warp matrix for ego-motion estimation and blur-matrix for latent image estimation [26, 42, 52, 48]. The world-to-sensor mapping in a narrow-angle system *sans* camera motion is $\mathbf{x} = (1/Z) \cdot K^n(\mathbf{X} + \mathbf{l_b})$ (which is obtained by imposing $R_t = I\ \forall t \in t_e^n$ in Eq. (2)). Relating the above equation with the mapping corresponding to a single pose-change (*i.e.*, $R_t = R\ \forall t \in t_e^n$) yields the pixel-mapping of a (homogeneous) coordinate $\mathbf{x}$ as

$$\mathbf{x}' = \lambda\Big(K^n R(K^n)^{-1}\mathbf{x} + \underbrace{\frac{1}{Z}K^n(I - R)\mathbf{l_c}}_{\text{center-of-rotation}} + \underbrace{\frac{1}{Z}K^n(I - R)\mathbf{l_b}}_{\text{baseline}}\Big),$$
$$(4)$$

where Z is the scene-depth corresponding to coordinate $\mathbf{x}$, and λ normalizes the third coordinate of $\mathbf{x}'$. (Full derivation

is provided in the Supplementary). Point spread function (PSF) at a spatial coordinate $\mathbf{x}$ is obtained by superimposing the pixel-mappings of $\mathbf{x}$ for all pose-changes undergone during the exposure time. Note that PSFs over spatial coordinates *completely* characterize motion blur (*i.e.*, motion blurred image is obtained by the space-variant convolution of PSFs and latent image) [48, 42]. An important insight from Eqs. (2)-(4) is that *PSF (and hence motion blur) in a DL set-up is depth-variant due to the baseline and COR, with its sensitivity increasing from farther to nearer scene-features (in addition to spatial variance)*. Wide-angle image can be represented akin to Eqs. (3) and (4) by enforcing $\mathbf{l_b} = \mathbf{0}$, and with a *different* MDF w^w and projection P^w.

3. A New Prior for Unconstrained DL-BMD

In this section, we first attempt to directly formulate a cost using Eqs. (3)-(4) for DL-BMD. Then we show that this approach is *untenable* for unconstrained DL set-ups, and warrants an additional prior.

The joint cost for DL-BMD is $L = L^n + L^w$:

$$L^k = \|\mathbf{A}^k \mathbf{w}^k - \mathbf{I}_B^k\|_2^2 + \lambda_1^k \|\mathbf{w}^k\|_1 + \lambda_2^k \|\nabla \mathbf{I}_C^k\|_1,$$
$$\text{where } \|\mathbf{A}^k \mathbf{w}^k - \mathbf{I}_B^k\|_2^2 = \|\mathbf{M}^k \mathbf{I}_C^k - \mathbf{I}_B^k\|_2^2. \quad (5)$$

where $k \in \{n, w\}$, $\mathbf{I}_C^k$ is the clean image, and $\mathbf{w}^k$ is the vectorized form of $w^k(p)$ (where p is an element of the discretized pose-space $\mathbb{P}^3$, named $\mathbb{P}_d^3$). The cost is derived as follows: For MDF $\mathbf{w}^k$, Eq. (3) enforces a linear relation via warp matrix $\mathbf{A}^k$, wherein its ith column contains the warped version of clean image $\mathbf{I}_C^k$, with the pose of $w^k(i)$ [48, 52], in accordance with Eq. (4). For clean image $\mathbf{I}_C^k$, Eq. (4) enforces a linear relation (*i.e.*, space-variant convolution) via PSF matrix $\mathbf{M}^k$, wherein its ith column contains the PSF corresponding to the ith coordinate. The term $\|\mathbf{w}^k\|_1$ enforces a prior on MDF that a 1D camera-path over time represents a sparse population in the 3D pose-space, and $\|\nabla \mathbf{I}_C^k\|_1$ enforces the total-variation image prior [31, 48, 4]. *Note that $\mathbf{A}^k$ and $\mathbf{M}^k$ are depth-dependent and are unique to DL set-up*, via baseline and COR in Eq. (4).

As discussed before, the estimated deblurred image-pair $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ must be related through scene-consistent disparities, *i.e.*, the narrow-angle camera must perceive the *same* scene-orientation, displaced by the baseline $\mathbf{l_b}$, as that by the wide-angle camera (*e.g.*, $\mathbf{I}_C^n = P^n(\mathbf{X} + \mathbf{l_b})$, if $\mathbf{I}_C^w = P^w(\mathbf{X})$). However, directly considering the DL-BMD cost for estimating $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ is *untenable*, as stated below:

Claim 1: There exist *multiple* valid solutions of deblurred image-pairs (or ill-posedness) for the DL-BMD cost (L in Eq. (5)) but that produce *scene-inconsistent disparities*.

<u>Proof:</u> A desired solution which minimizes Eq. (5) is the one involved in the blurring process (Eq. (3)), which we refer to as the true image-pair $\{P^n(\mathbf{X} + \mathbf{l_b}), P^w(\mathbf{X})\}$ and true MDFs $\{w^n(p), w^w(p)\}$. Though not characterizing the

(a) GT MDF (b) W/ Prior (c) W/o Prior (d) Image patches
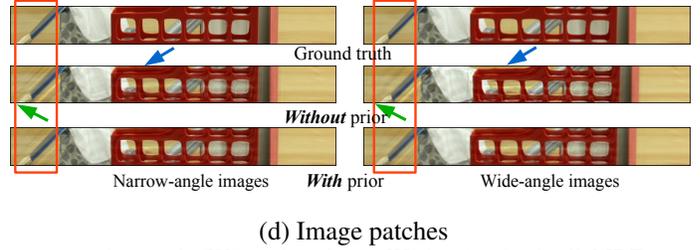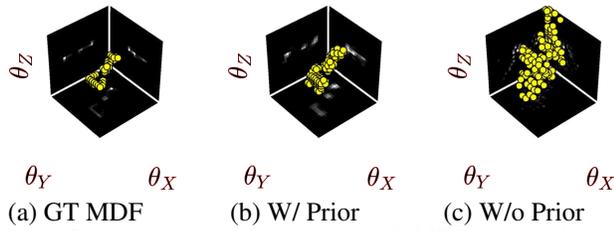
Figure 1. Effect of the proposed prior: (a-d) MDFs and deblurred image patches with (W/) and without (W/o) prior (with all MDFs centroid-aligned with the ground truth (GT) $\mathbf{w}^n$ to align left-images). MDF estimate of the prior-less case has a random offset (Fig. (c)) and the corresponding deblurred image clearly reveals *scene-inconsistent* disparities (Fig. (d)). Also, the deblurred image in the prior-less case exhibits considerable ringing artifacts and residual blur (Fig. (d)). In contrast, the addition of our proposed DL prior successfully curbs the pose ambiguity and improves the MDF accuracy (Fig. (b)) and produces better deblurring quality (Fig. (d)).

blur process per se, Eq. (3) can be equivalently written as

$$\mathbf{I}_B^n = \sum_p w^n(p) P^n \big(R_p R_n^{-1} R_n (\underbrace{\mathbf{X}}_{true} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b} \big),$$

$$= \sum_p w^n(p) P^n \big(R_p R_n^{-1} (\underbrace{R_n(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c}}_{apparent} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b} \big),$$

(6)

where the new scene-orientation of narrow-angle lens is $R_n(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c}$, where $R_n \neq I$. The quantity R_n has the effect of shifting all the true poses undergone by the camera $(R_p, p \in \mathbb{P}_d^3)$ by an offset of R_n^{-1}, which in turn produces an MDF that is a shifted version of the *true* MDF (and hence the MDF-sparsity cost remains the same). Consequently, a new solution according to Eq. (6) is the image-pair $\{P^n(R_n(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}), P^w(\mathbf{X})\}$, which clearly fails the criterion for scene-consistent disparities (*i.e.*, the narrow-angle camera perceives a *different* scene-orientation). Also, as the new narrow-angle image is a warped version of the true narrow-angle image, it adheres to the TV prior, and therefore the new solution minimizes L^n. The cost L^w remains the same (as the wide-angle image or MDF incurs *no* change). Resultantly, the same solution minimizes L, which concludes the proof. ∎
A similar ambiguity also arises for the wide-angle case. This is obtained from Eq. (6) by enforcing $\mathbf{l_b} = \mathbf{0}$ and replacing P^n by P^w. As the costs L^n and L^w (in Eq. (5)) are *independent*, the pose R_n need *not* be equal to that of wide-angle (R_w). For unequal R_n and R_w, the resultant image-pair becomes $\{P^n(R_n(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b}), P^w(R_w(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c})\}$. Following the similar steps in the proof, we can show that the resultant solution minimizes L, though the image-pairs produce *scene-inconsistent* disparities.

We attempt to provide some insights on the effect of ill-posedness. Consider the case of a positive inplane rotation ambiguity, with COR at the optical center. Figure 2(a) shows three image coordinates $\{A, B, C\}$ with *identical* scene-depths (*i.e.*, the same disparities). Fig. 2(b) considers the rotational ambiguity, *i.e.* the coordinates $\{A, B, C\}$


(a) Desired solution (b) An ill-posed solution
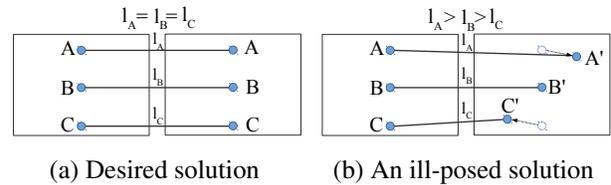
Figure 2. $\{A, B, C\}$ in Fig. (a) correspond to scene-features at the same depth (*i.e.*, *identical* disparities). Fig. (b) considers an inplane rotational ambiguity, wherein $\{A, B, C\}$ translates to $\{A', B', C'\}$ which clearly leads to *inconsistent* disparities.

are mapped to $\{A', B', C'\}$, respectively. It is evident from Fig. 2(b) that, relative to the scene-feature of B, A's scene-feature appears to be farther and C's scene-feature appears to be nearer, even though all the scene features have identical depths in the world system.

Note that the ill-posedness *exists* irrespective of the exposure time being identical or different. Moreover, the inconsistent deblurred image-pair shares all the issues associated with the classical problem of stereo rectification [21, 50] that deals with *physical* misalignment of cameras. These methods work by estimating a pair of homography for rectification [50, 8]. However, the ambiguity in DL-BMD is different, in that it necessitates *depth-variant* transformation due to baseline and arbitrary COR (Eq. (4)).

We tackle the ill-posedness *within* our deblurring method, by employing a judiciously derived prior. For this, we assume that there exists an overlap between exposure times of different cameras. A DL set-up that violates this assumption has to incur significant ghosting artifacts, and is hence *not* preferred [29]. Note that our assumption is generic as compared to that of the *complete* exposure-time overlap in the only-existing DL-BMD method [51].

Our prior is motivated by the previous discussion, in that the deblurred image-pair will be consistent if $R_n = R_w$. For identical exposure time, this criterion requires that both the MDFs completely intersect over the pose-space. For

overlapping exposure time, both MDFs must intersect over the shared poses. Hence, we introduce a DL prior of the form $\|\mathbf{w}^n - \mathbf{w}^w\|_2$. Intuitively, the prior functions as follows: The DL-BMD cost can admit MDF-pairs with significant relative drifts, which severely disrupt scene-consistent disparities (*e.g.*, see Figs. 1(c,d)). However, these solutions are not favoured with the inclusion of the prior because it enforces the resultant cost to increase with relative drifts (*e.g.*, see Figs. 1(b,d)).

The proposed DL prior has several desirable properties: As shown in [31, 52, 6, 9], the biconvexity property (*i.e.*, the BMD cost is *convex* with respect to MDF for a given clean image, and vice-versa) guarantees convergence via alternating minimization. Our final cost has this property.

Claim 2: The DL-BMD cost L (Eq. (5)) is biconvex with respect to image-pair $\{\mathbf{I}_C^n, \mathbf{I}_C^w\}$ and MDF-pair $\{\mathbf{w}^n, \mathbf{w}^w\}$. The DL prior is convex, and when added to the cost L retains the biconvexity property. (Proof is provided in the supplementary material.) Also, our prior serves to impart reinforcement between the dual images (through MDFs), which Eq. (5) does *not* possess (as L_n and L_w are independent). It aids in ego-motion estimation, which in turn leads to improved deblurring (*e.g.*, see Fig. 1(d)). Also, the prior allows for efficient LASSO optimization (as we shall see in Section 4.2).

4. A Practical algorithm for DL-BMD

In this section, we propose a practical DL-BMD algorithm for unconstrained camera settings and arbitrary COR (a first of its kind), based on the proposed model and DL prior (Secs. 2-3). We show that a multi-camera BMD problem can be divided into subproblems (with the same dimension as that of normal camera BMD) while enforcing the DL prior and convexity property.

Our method proceeds in a scale-space manner to handle large blurs [26, 52, 48, 6]. We employ alternating minimization (AM) for depth, COR, MDF and latent image, in that order. The convergence of AM is supported by Sec. 3, in that resolving the ill-posedness enforces scene-consistent image-pair, which in turn produces consistent depth and COR [12]. As 'depth from stereo' is a well-studied problem, we selected an off-the-shelf algorithm for depth estimation [20] (owing to its good trade-off between accuracy and speed [19, 38]).

4.1. Center-of-Rotation Estimation

To estimate COR, we consider a cost which is the least squares error between blurred images and synthesized blurry images using the blur model (via Eqs. (3)-(4)) and current estimates of other unknowns. We frame the cost in the gradient domain of the images to improve the condition number [6]. In order to ensure that all regions of the image constrain COR, the image is split into multiple bins and

thresholding is done separately for each bin. The optimization for COR is given as $\tilde{l}_c = \arg\min_{\mathbf{l_c}} \left(L_{\mathbf{l_c}}^w + L_{\mathbf{l_c}}^n \right)$:

$$L_{\mathbf{l_c}}^k = \|g(\mathbf{I}_B^k) - g\left(\sum_p \tilde{w}^k(p) P^k(\tilde{\mathbf{I}}_C^k, \tilde{\mathbf{Z}}, \mathbf{l_c}) \right)\|_2, \quad (7)$$

where $k \in \{w, n\}$, $g(\cdot)$ produces the first and second-order gradients, and the symbol '$\sim$' denotes the current estimates. A trust region reflective algorithm [7] is used for optimizing Eq. (7), which is initialized with the previous COR estimate. For the first scale and first iteration, we initialize the latent images as the corresponding shock-filtered blurred images, MDFs as Kronecker delta, and COR at the optical center.

4.2. Divide Strategy for MDFs and Images

Jointly estimating multiple MDFs or images is computationally inefficient, as the optimization dimension scales-up linearly with each additional camera input. To this end, we decompose the DL-BMD cost with prior, such that convexity is preserved and the optimization dimension remains at par with that of normal camera, irrespective of the number of cameras. The MDF and image estimation are given by

$$\arg\min_{\mathbf{w}^n} \|\tilde{\mathbf{A}}^n \mathbf{w}^n - \mathbf{I}_B^n\|_2^2 + \alpha\|\mathbf{w}^n - \tilde{\mathbf{w}}^w\|_2^2 : \|\mathbf{w}^n\|_1 \leq \lambda_1'^n,$$

$$\arg\min_{\mathbf{I}_C^n} \|\tilde{\mathbf{M}}^n \mathbf{I}_C^n - \mathbf{I}_B^n\|_2^2 + \lambda_2\|\nabla\mathbf{I}_C^n\|_1, \quad (8)$$

where we have included the DL prior within the objective, but separated out the MDF-sparsity prior as a constraint. Using Claim 2, we can show that individual optimizations in Eq. (8) are convex. Further, though nontrivial, MDF estimation with the DL prior (in Eq. (8)) has an *equivalent* LASSO form $\arg\min_{\mathbf{w}^n} \|\mathbf{K}\mathbf{w}^n - \mathbf{b}\|_2^2 : \|\mathbf{w}^n\|_1 \leq \lambda_1'^n$, such that (proofs are provided in the supplementary)

$$\mathbf{K} = \tilde{\mathbf{A}}^{n^T}\tilde{\mathbf{A}}^n + \alpha I, \text{ and } \mathbf{b} = \tilde{\mathbf{A}}^{n^T}\mathbf{I}_B^n + \alpha\tilde{\mathbf{w}}^n. \quad (9)$$

A similar formulation as that of Eqs. (8)-(9) applies to the other camera as well. We optimized for MDFs using the standard LASSO solver [45] (following [48, 6]). Also, our divide strategy converts the latent image estimation to the classic problem of TV-deblurring [4] (the only difference is that $\tilde{\mathbf{M}}^n$ is now in accordance with DL-model), which has excellent convergence and efficient solvers [31]. As image estimators are independent, they can be parallelized for efficiency. These are made possible by our decomposition of the DL-BMD problem while enforcing the DL prior.

5. Analysis and Discussions

In this section, we indicate the generalizability of our work to diverse camera set-ups. Then, we analyse the effect of our prior and COR, and discuss further implications.

PSNR (dB)	Blur	W/o Prior W/o COR	W/o Prior W/ COR	W/ Prior W/o COR	W/ prior W/ COR
Image	22.39	25.69	26.59	27.28	28.88
Depth	28.33	23.35	**23.59**	29.12	**30.52**

Table 1. Quantitative results of our method with and without the DL prior and COR. In particular, our DL prior reduces the ill-posedness by a good margin (*i.e.*, by 7 dB, as indicated in bold).

Generalizability: Our theory and method directly apply to DL cameras with entirely different settings. Second, they hold well for *identical* cameras ($f^n = f^w$) or camera arrays (multiple l_b), wherein exposures are different ($\mathbf{w}^n \neq \mathbf{w}^w$ or $\mathbf{w}^n = \mathbf{w}^w$) or identical ($\mathbf{w}^n = \mathbf{w}^w$). Third, they generalize to the mature normal camera methods ($\mathbf{l_b} = \mathbf{l_c} = \mathbf{0}$ and $\mathbf{w}^n = \mathbf{w}^w$) [26, 52, 48]. Based on the previous discussions, we make the following remarks.
Remark 1: The motion blur model of the methods [26, 52, 48] admits *only* a depth invariant model, whereas motion blur in a DL set-up warrants a depth variant model.
Remark 2: The blur model of the methods [26, 52, 48] modulate the baseline with camera poses, but it must be independent for a DL set-up (for scene-consistent disparities).
Remark 3: The methods [26, 48, 52] also admit the ill-posedness that disrupts scene-consistent disparities.
(Proofs with illustrations are provided in supplementary).
Effectiveness of the DL prior and COR: Table 1 summarizes the PSNR results for image/depth (averaged over five examples) by ablating the DL prior and COR estimator. For creating synthetic dataset, exposure overlap and COR are randomly sampled from 10 to 100% and −30 to 30 cm cube, respectively. The unconstrained set-up we employed is narrow- and wide-FOV pair, with $f^n = 52$ mm, $f^w = 26$ mm, and the former having twice the resolution (as in `Samsung S9+`). Observe that for the prior-less case the depth information gets significantly corrupted (*i.e.*, PSNR drops by 7 dB!). This underlines the importance of resolving the pose-ambiguity in dual-lens BMD. Further, the deblurring performance also drops by 2.3 dB in the prior-less case, possibly be due to the loss of reinforcement between the narrow- and wide-angle costs (as discussed earlier). Further, the table reveals that both image and depth accuracies deteriorate when COR issue is *not* addressed, *i.e.*, image and depth PSNRs drop by 1.6 and 1.3 dB, respectively.
Implications: Our method can *seamlessly* address partial and full exposure-overlaps ([14, 29, 47, 30, 24]), *without* any modifications. Further, rolling shutter effects are pertinent problems in well-lit scenarios, but they have *not* been addressed for unconstrained cameras. They also require a 'homography-like' warping (such as Eq. (4)), admit the *same* ill-posedness, and hence necessitate an analogous prior. For deep learning, Eqs. (2)-(4) can potentially aid in generating datasets (which are currently non-existent).
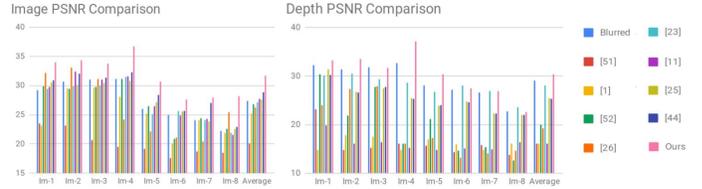
Figure 3. Quantitative evaluations using objective measure (PSNR). Our method performs competitively against the state-of-the-art, and produces the least depth errors.
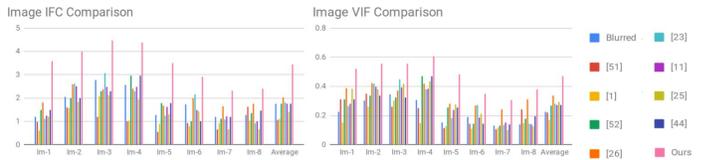

Figure 4. Quantitative evaluations using subjective measures (IFC, VIF). Our method performs deblurring with the best aesthetics.

6. Experimental Results

In this section, we extensively evaluate our proposed method on both synthetic and real examples.
Comparison Methods: We considered [26, 52] to represent normal camera BMD. For computational cameras, we considered state-of-the-art stereo BMD [51] and light field BMD [23]. For depth-aware case, we considered the single-image BMD [11] and multi-image method [1]. For deep learning, we considered [44, 25] which represent recurrent and autoencoder networks, respectively. Note that the publicly available code for [5, 40] require as input 4D light field, whereas the codes for [27, 18] are not available.
Metrics: For quantitative evaluation of image, we employ PSNR, IFC [36], and VIF [35]. We have selected IFC and VIF because they are shown to be the best metrics for subjective evaluation of BMD [17]. For qualitative evaluation, we provide the narrow-FOV image and (normalized) depth estimated from deblurred image-pair or by algorithms [11, 1]. Due to space constraints, we consider all methods for one example and provide sparse comparisons for others. Nevertheless, our supplementary covers all methods.
Quantitative Evaluation: Figures 3-4 provide objective and subjective measures for different methods. First of all, both the measures of the state-of-the-art DL-BMD [51] clearly reveal its high sensitivity, when it deviates from the assumptions of synchronized and identical cameras, and layered depth scenes. This once again emphasizes the need for an unconstrained DL-BMD method. For normal camera methods [26, 52], there is a perceivable drop in the depth performance (due to Remarks 2-3), which clearly suggests their inadequacy in DL set-up. While the inferior depth performance of [1] can be attributed to its assumption of lay-
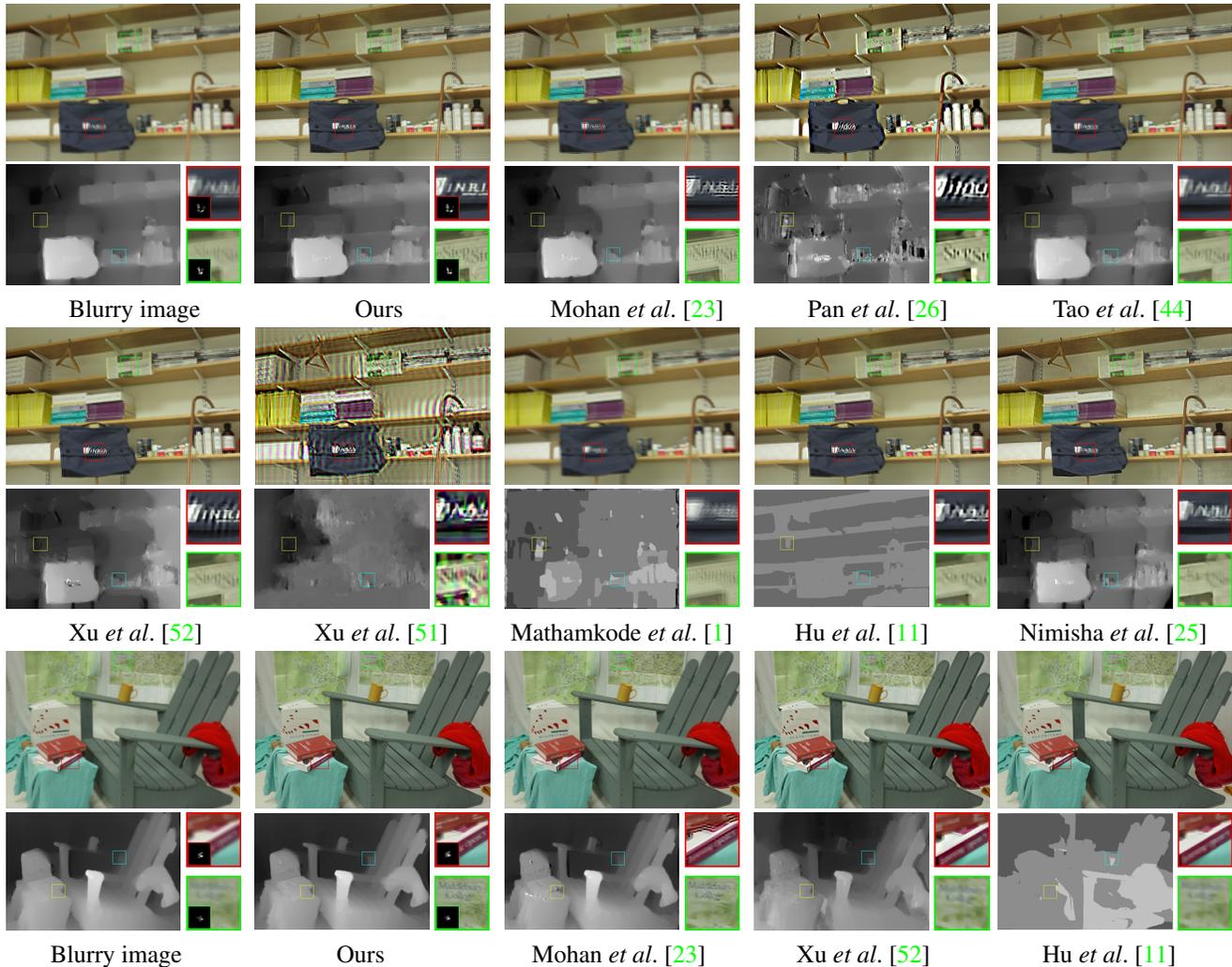
Figure 5. Synthetic experiments: The method of [51, 11, 1] exhibits severe ringing artifacts and inaccurate depth estimates. The results of [26, 52] amply underline the shortcomings of normal camera models. As compared to deep learning [44, 25] and light field BMD [23], our method retrieves distinct textual information. Also, we compare depth- and space-variant GT and estimated PSFs (inset patches of blurry and our results).

ered depth, for [11], it can also be due to its single image restriction. As compared to our method, light field BMD [23] is not quite successful (*i.e.*, image/depth PSNR is less by 2.37/4.47 dB). This can be attributed to its lens effect and assumption of synchronized and identical camera settings. Our method outperforms deep learning methods [25, 44] by 3.50 dB and 2.72 dB for image and 4.39 dB and 4.36 dB for depth, respectively. Based on the claims of [25, 44] that they generalize well for real-captured images, this performance degradation could be possibly due to the unique characteristics of unconstrained DL blur.

Qualitative Evaluation: Figures 5-6 provide visual results for synthetic [33] and real experiments. We wish to highlight that ringing artifacts in deblurring are mainly caused by ego-motion error, which can be either due to inaccu-

rate blur/ego-motion model or ineffectiveness of optimization. It can be seen that depth estimation is *also* sensitive to ringing artifacts; one reason could be that ringing deteriorates the feature matches required for depth estimation. The deblurred images of [51, 1] exhibit severe ringing artifacts (possibly due to the assumptions on scene and ego-motion and capture settings). Also, note that [11] produces erroneous layered-depth estimates (*e.g.*, nearer depths appear to be farther, as in Fig. 6, first row, chandelier). This is due to its sole restriction to single image cues for depth sensing. The results of [23, 26, 52] amply demonstrate the inadequacy of light field and single-lens BMD in the dual-lens setup, where the deblurring is *not* uniform over different depth levels (*e.g.*, in Fig. 5, fifth row, the closer books and farther windows are *not* simultaneously accounted for) and
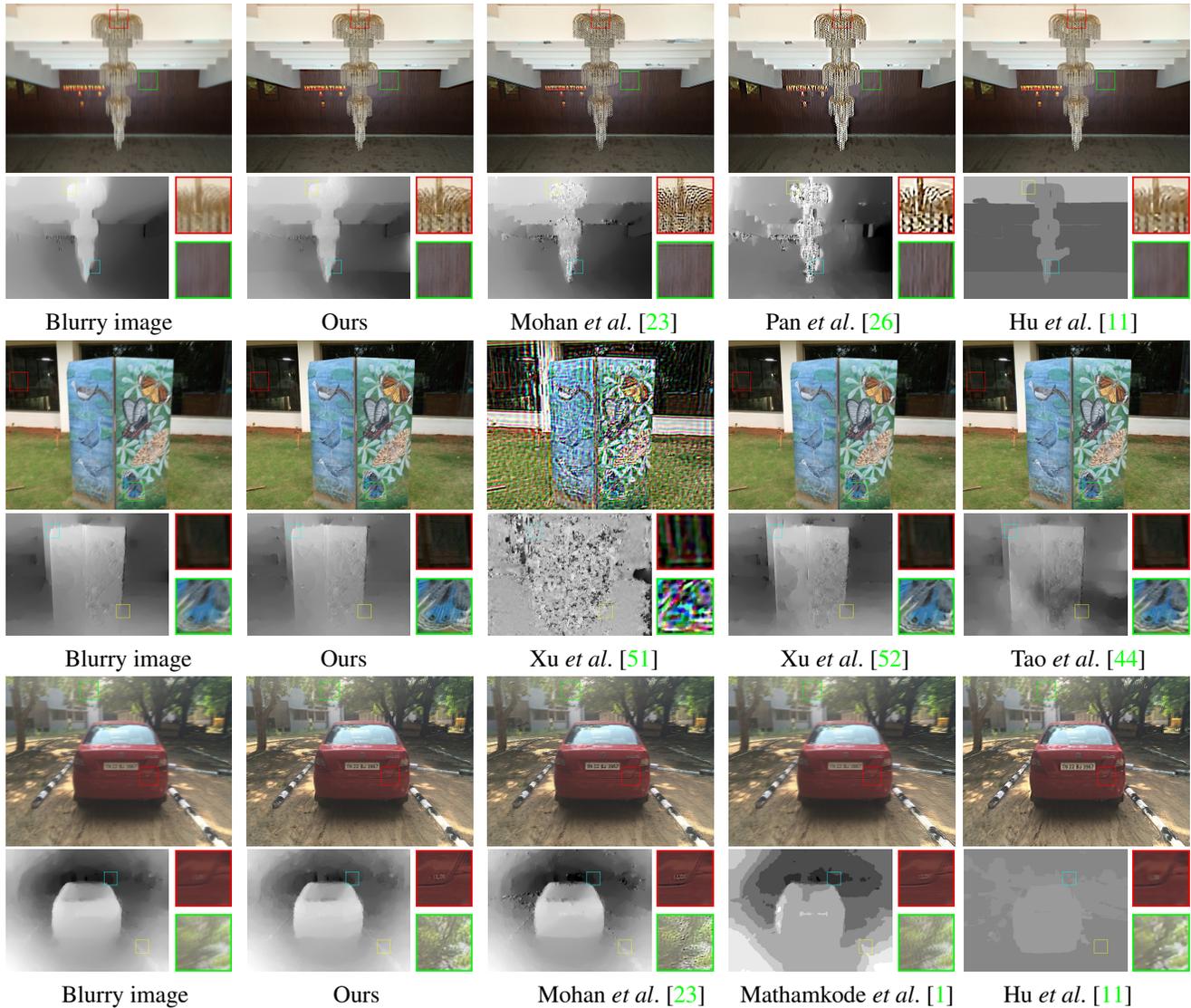
Figure 6. Real experiments: (first row - indoor scene, second and third row - outdoor scenes). Our method is able to recover finer features at different depth ranges as compared to the competing methods, and is able to faithfully preserve the depth information.

exhibits perceivable ringing artifacts, (*e.g.*, in Fig. 6, first row, over the chandelier). The visual results of deep learning methods [25, 44] once again prove that they are inadequate to deal with DL blur. When compared with the competing methods on all the examples, it is evident that our DL deblurring method consistently accounts for features at different depths, produces lesser ringing artifacts, and faithfully preserves consistent depth information. (Please refer to our supplementary for further analysis and examples.)

7. Conclusions

In this paper, we addressed the problem of blind motion deblurring for unconstrained dual-camera set-ups. Our algorithm allows for any arbitrary COR in the blurring process and is incorporated in the optimization pipeline. We revealed an inherent ambiguity in the BMD problem which hampers the scene-consistent depth cues embedded in the image-pair. Towards this end, we introduced a convex and computationally efficient prior. We showed the efficacy of the proposed prior which enforces scene consistent disparities, leading to improved deblurring. Comprehensive comparisons with existing state-of-the-art methods amply demonstrate the superiority and need of our method. As an increasing number of modern cameras are employing dual-lens configurations, our theory and method will be very relevant for steering further research in this field.

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. 2, 6, 7, 8

[2] Michel Bätz, Thomas Richter, Jens-Uwe Garbas, Anton Papst, Jürgen Seiler, and André Kaup. High dynamic range video reconstruction from a stereo camera setup. *Signal Processing: Image Communication*, 29(2):191–202, 2014. 1

[3] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, pages 25–36. Springer, 2004. 2

[4] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE transactions on Image Processing*, 7(3):370–375, 1998. 3, 5

[5] Paramanand Chandramouli, Meiguang Jin, Daniele Perrone, and Paolo Favaro. Plenoptic image motion deblurring. *IEEE Transactions on Image Processing*, 27(4):1723–1734, 2018. 1, 2, 6

[6] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *ACM Transactions on Graphics (TOG)*, 28(5):145, 2009. 1, 2, 5

[7] Thomas F Coleman and Yuying Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2):418–445, 1996. 5

[8] Andrea Fusiello, Emanuele Trucco, and Alessandro Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 4

[9] Ankit Gupta, Neel Joshi, C Lawrence Zitnick, Michael Cohen, and Brian Curless. Single image deblurring using motion density functions. In *European Conference on Computer Vision (ECCV)*, pages 171–184. Springer, 2010. 1, 5

[10] Sung Hee Park and Marc Levoy. Gyro-based multi-image deconvolution for removing handshake blur. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3366–3373, 2014. 1

[11] 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. 2, 6, 7, 8

[12] Zhe Hu, Lu Yuan, Stephen Lin, and Ming-Hsuan Yang. Image deblurring using smartphone inertial sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1855–1864, 2016. 1, 2, 3, 5

[13] Ganesh Iyer, J Krishna Murthy, Gunshi Gupta, Madhava Krishna, and Liam Paull. Geometric consistency for self-supervised end-to-end visual odometry. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 267–275, 2018. 1

[14] Daniel S Jeon, Seung-Hwan Baek, Inchang Choi, and Min H Kim. Enhancing the spatial resolution of stereo images using a parallax prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1721–1730, 2018. 1, 6

[15] Neel Joshi, Sing Bing Kang, C Lawrence Zitnick, and Richard Szeliski. Image deblurring using inertial measurement sensors. *ACM Transactions on Graphics (TOG)*, 29(4):30, 2010. 1

[16] 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. 2

[17] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1701–1709, 2016. 6

[18] Dongwoo Lee, Haesol Park, In Kyu Park, and Kyoung Mu Lee. Joint blind motion deblurring and depth estimation of light field. In *European Conference on Computer Vision (ECCV)*, pages 288–303. Springer, 2018. 2, 3, 6

[19] Bing Li, Chia-Wen Lin, Boxin Shi, Tiejun Huang, Wen Gao, and C-C Jay Kuo. Depth-aware stereo video retargeting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, 2018. 5

[20] Ce Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. 5

[21] Charles Loop and Zhengyou Zhang. Computing rectifying homographies for stereo vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 125–131, 1999. 4

[22] Pei-Ying Lu, Tz-Huan Huang, Meng-Sung Wu, Yi-Ting Cheng, and Yung-Yu Chuang. High dynamic range image reconstruction from hand-held cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 509–516, 2009. 1

[23] 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, 2, 6, 7, 8

[24] Jiawei Mo and Junaed Sattar. Dsvo: Direct stereo visual odometry. *arXiv preprint arXiv:1810.03963*, 2018. 1, 6

[25] 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. 6, 7, 8

[26] 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, 2, 3, 5, 6, 7, 8

[27] Liyuan Pan, Yuchao Dai, and Miaomiao Liu. Single image deblurring and camera motion estimation with depth map. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2116–2125, 2019. 2, 6

[28] Liyuan Pan, Yuchao Dai, Miaomiao Liu, and Fatih Porikli. Simultaneous stereo video deblurring and scene flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6987–6996, 2017. 2

[29] Won-Jae Park, Seo-Won Ji, Seok-Jae Kang, Seung-Won Jung, and Sung-Jea Ko. Stereo vision-based high dynamic

range imaging using differently-exposed image pair. *Sensors*, 17(7):1473, 2017. 1, 2, 4, 6

[30] NF Pashchenko, KS Zipa, and AV Ignatenko. An algorithm for the visualization of stereo images simultaneously captured with different exposures. *Programming and Computer Software*, 43(4):250–257, 2017. 1, 2, 6

[31] Daniele Perrone and Paolo Favaro. Total variation blind deconvolution: The devil is in the details. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2909–2916, 2014. 2, 3, 5

[32] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics (TOG)*, 23(3):664–672, 2004. 1

[33] 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. 7

[34] Anita Sellent, Carsten Rother, and Stefan Roth. Stereo video deblurring. In *European Conference on Computer Vision (ECCV)*, pages 558–575. Springer, 2016. 2

[35] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on Image Processing (TIP)*, 15(2):430–444, 2006. 6

[36] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing (TIP)*, 14(12):2117–2128, 2005. 6

[37] Xiaoyong Shen, Hongyun Gao, Xin Tao, Chao Zhou, and Jiaya Jia. High-quality correspondence and segmentation estimation for dual-lens smart-phone portraits. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3277–3286, 2017. 1

[38] Kuang-Tsu Shih and Homer H Chen. Generating high-resolution image and depth map using a camera array with mixed focal lengths. *IEEE Transactions on Computational Imaging*, 5(1):68–81, 2018. 5

[39] Ondrej Sindelar and Filip Sroubek. Image deblurring in smartphone devices using built-in inertial measurement sensors. *Journal of Electronic Imaging*, 22(1):011003, 2013. 1

[40] Pratul P. Srinivasan, Ren Ng, and Ravi Ramamoorthi. Light field blind motion deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3958–3966, 2017. 1, 2, 3, 6

[41] Filip Sroubek and Peyman Milanfar. Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Transactions on Image processing (TIP)*, 21(4):1687–1700, 2012. 1, 2

[42] Shuochen Su and Wolfgang Heidrich. Rolling shutter motion deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1529–1537, 2015. 2, 3

[43] Ning Sun, Hassan Mansour, and Rabab Ward. Hdr image construction from multi-exposed stereo ldr images. In *IEEE International Conference on Image Processing*, pages 2973–2976, 2010. 1, 2

[44] 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. 6, 7, 8

[45] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. 5

[46] F. Wang, T. Li, and Y. Li. Dual deblurring leveraged by image matching. In *IEEE International Conference on Image Processing*, pages 567–571, 2013. 1

[47] Jian Wang, Tianfan Xue, Jonathan T Barron, and Jiawen Chen. Stereoscopic dark flash for low-light photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, 2019. 1, 2, 6

[48] 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, 2, 3, 5, 6

[49] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics (TOG)*, 24(3):765–776, 2005. 2

[50] Ruichao Xiao, Wenxiu Sun, Jiahao Pang, Qiong Yan, and Jimmy Ren. Dsr: Direct self-rectification for uncalibrated dual-lens cameras. In *IEEE International Conference on 3D Vision (3DV)*, pages 561–569, 2018. 4

[51] Li Xu and Jiaya Jia. Depth-aware motion deblurring. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2012. 1, 2, 4, 6, 7, 8

[52] 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, 2, 3, 5, 6, 7, 8

[53] Xiaoyun Yuan, Ziwei Xu, Haoqian Wang, Yebin Liu, and Lu Fang. Cascaded image deblurring with combined image prior. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 16–20, 2017. 1

[54] Cha Zhang and Tsuhan Chen. A self-reconfigurable camera array. In *Eurographics Conference on Rendering Techniques*, pages 243–254. Eurographics Association, 2004. 2

[55] Haichao Zhang, David Wipf, and Yanning Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1051–1058, 2013. 1

[56] Li Zhang, Alok Deshpande, and Xin Chen. Denoising vs. deblurring: Hdr imaging techniques using moving cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 522–529, 2010. 1

[57] Xiang Zhu, Filip Šroubek, and Peyman Milanfar. Deconvolving psfs for a better motion deblurring using multiple images. In *European Conference on Computer Vision (ECCV)*, pages 636–647. Springer, 2012. 1

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

$$\mathbf{X}'' = \mathbf{X} + \mathbf{l_b}, \tag{S1}$$

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

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

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

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{l_c}) + \mathbf{l_c} + \mathbf{l_b} = R(\mathbf{X} + \mathbf{l_b}) + (I - R)\mathbf{l_c} + (I - R)\mathbf{l_b}. \tag{S3}$$

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

$$\mathbf{X}' = ZR(K^n)^{-1}\mathbf{x} + (I - R)\mathbf{l_c} + (I - R)\mathbf{l_b}. \tag{S4}$$

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

$$\begin{aligned}
\mathbf{x}' = K^n \frac{\mathbf{X}'}{Z'} &= \frac{Z}{Z'} K^n R(K^n)^{-1}\mathbf{x} + \frac{1}{Z'} K^n (I - R)\mathbf{l_c} + \frac{1}{Z'} K^n (I - R)\mathbf{l_b}, \\
&= \frac{Z}{Z'}\Big(K^n R(K^n)^{-1}\mathbf{x} + \frac{1}{Z} K^n (I - R)\mathbf{l_c} + \frac{1}{Z} K^n (I - R)\mathbf{l_b}\Big).
\end{aligned} \tag{S5}$$

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

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

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

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

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

$$\begin{aligned}
L_{\mathbf{w}^n}^n\big(\gamma \mathbf{I_1}_C^n + (1-\gamma)\mathbf{I_2}_C^n\big) &\leq \gamma L_{\mathbf{w}^n}^n(\mathbf{I_1}_C^n) + (1-\gamma)L_{\mathbf{w}^n}^n(\mathbf{I_2}_C^n) \quad \forall\{\mathbf{I_1}_C^n, \mathbf{I_2}_C^n, \mathbf{w}^n\}, \; \forall\gamma \in [0,1]; \\
L_{\mathbf{I}_C^n}^n\big(\gamma \mathbf{w_1}^n + (1-\gamma)\mathbf{w_2}^n\big) &\leq \gamma L_{\mathbf{I}_C^n}^n(\mathbf{w_1}^n) + (1-\gamma)L_{\mathbf{I}_C^n}^n(\mathbf{w_2}^n) \quad \forall\{\mathbf{w_1}^n, \mathbf{w_2}^n, \mathbf{I}_C^n\}, \; \forall\gamma \in [0,1].
\end{aligned} \tag{S8}$$

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

$$\begin{aligned}
L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n\big(\gamma\{\mathbf{I_1}_C^n, \mathbf{I_1}_C^w\} + (1-\gamma)\{\mathbf{I_2}_C^n, \mathbf{I_2}_C^w\}\big) &\leq \gamma L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n(\{\mathbf{I_1}_C^n, \mathbf{I_1}_C^w\}) + (1-\gamma)L_{\{\mathbf{w}^n, \mathbf{w}^w\}}^n(\{\mathbf{I_2}_C^n, \mathbf{I_2}_C^w\}); \\
L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n\big(\gamma\{\mathbf{w_1}^n, \mathbf{w_1}^w\} + (1-\gamma)\{\mathbf{w_2}^n, \mathbf{w_2}^w\}\big) &\leq \gamma L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n(\{\mathbf{w_1}^n, \mathbf{w_1}^w\}) + (1-\gamma)L_{\{\mathbf{I}_C^n, \mathbf{I}_C^w\}}^n(\{\mathbf{w_2}^n, \mathbf{w_2}^w\}).
\end{aligned} \tag{S9}$$

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

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

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.

<u>Proof:</u> MDF optimization problem is given as (Eq. (8)):

$$\tilde{\mathbf{w}}^n = \arg\min_{\mathbf{w}^n} \underbrace{\|\tilde{\mathbf{A}}^n \mathbf{w}^n - \mathbf{I_B}^n\|_2^2 + \alpha\|\mathbf{w}^n - \tilde{\mathbf{w}}^w\|_2^2}_{G} \; : \|\mathbf{w}^n\|_1 \leq \lambda_1'^n, \tag{S11}$$

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*

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

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.*,

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

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 famework. 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.

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

$$\hat{\mathbf{x}}' = \lambda K \hat{R} K^{-1} \mathbf{x} \tag{S14}$$

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).

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

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

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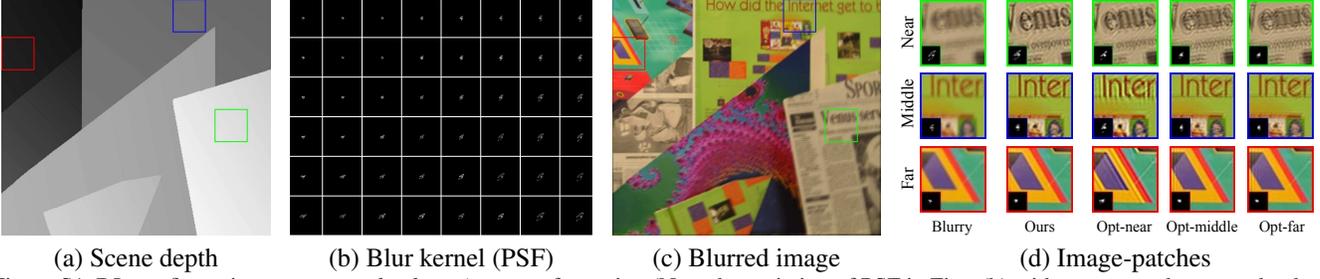
Figure S1. DL configuration warrants a *depth-variant* transformation (Note the variation of PSF in Fig. (b) with respect to the scene depth in Fig. (a)). As the single-lens motion blur model is *depth-invariant*, the model optimized for a fixed depth can fail for other depths, leading to *ineffective*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


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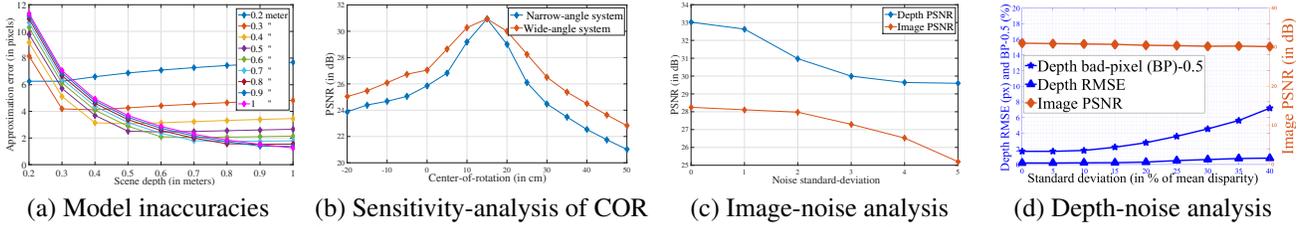
Figure S2. Analysis: (a) Model inaccuracies of the homography model. (b) Sensitivity of COR: Both narrow-angle and wide-angle configurations are very sensitive to COR, with the former exhibiting relatively more sensitivity. (c-d) Effect of image and depth noise.

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 $\hat{\mathbf{I}}_B^n = (1/t_e^n) \int_{t_e^n} P^n\big(\hat{R}_t(\mathbf{X} + \mathbf{l_b})\big)\, dt$ (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

(a) Wide-angle blurred   (b) True MDF   (c) MDF *with* prior   (d) MDF *without* prior

(e) Narrow-angle images   (f) Wide-angle images   (g) Wide-angle image-patches

(1) Ground truth

(2) *With* Prior

(3) *Without* Prior
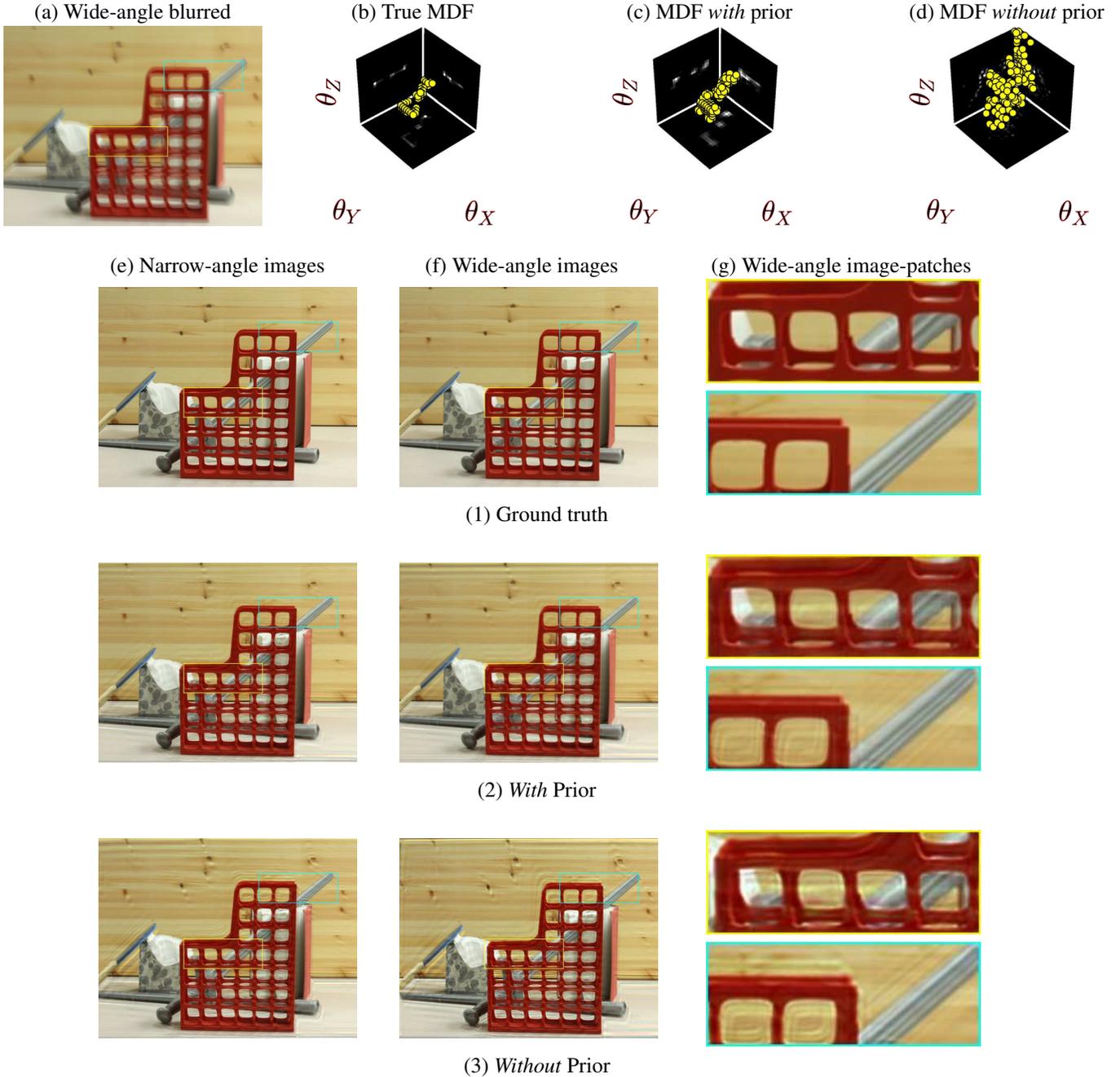
Figure S3. Effect of the proposed prior: The MDF estimate and the deblurred image-patches of prior-less case clearly show a significant rotational ambiguity (Figs. [g,3] and d). Also, the deblurred image in the prior-less case exhibits considerable ringing artifacts and residual blur (Figs. [e-f,3]), 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 *et al.* [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 ith 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
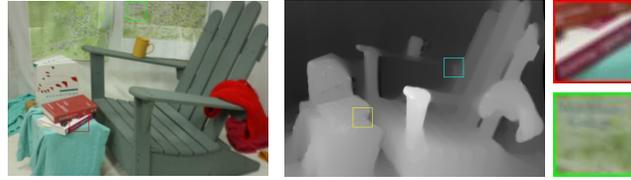
(a) Blurry image
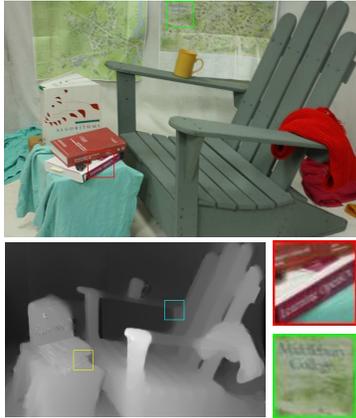
(b) Ours   (c) Mohan *et al.* [9]   (d) Pan *et al.* [12]
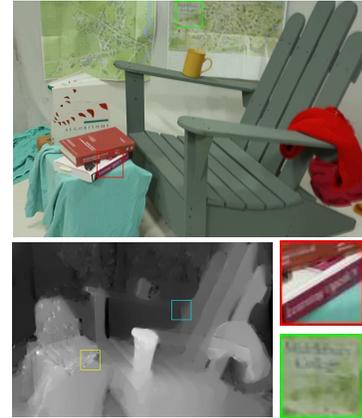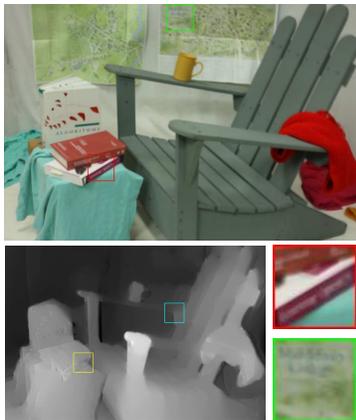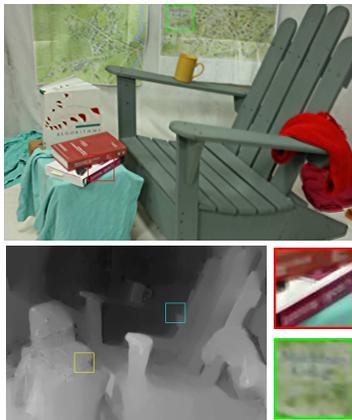
(e) Tao *et al.* [15]   (f) Xu *et al.* [18]   (g) Xu *et al.* [17]
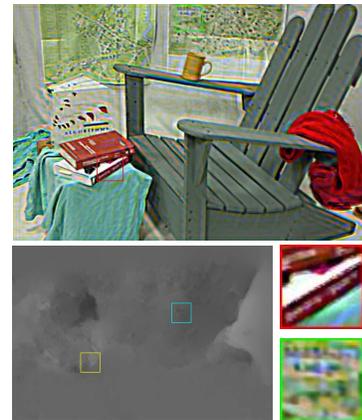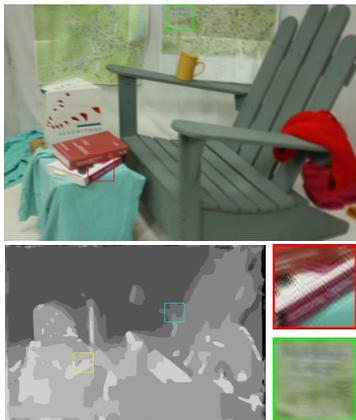
(e) Mathamkode *et al.* [1]   (f) Hu *et al.* [5]   (g) Nimisha *et al.* [11]

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.
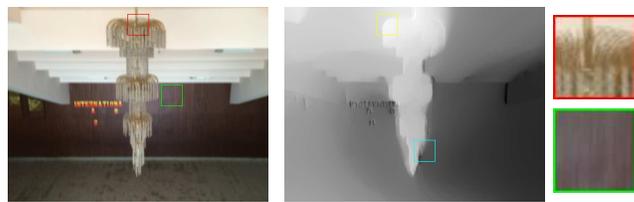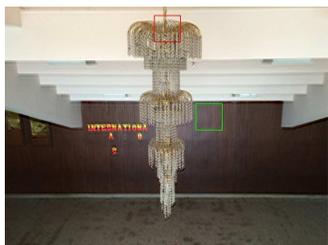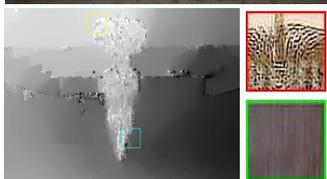
(a) Blurry image

(b) Ours   (c) Mohan *et al.* [9]   (d) Pan *et al.* [12]
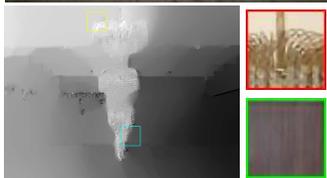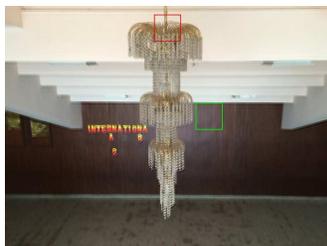
(e) Tao *et al.* [15]   (f) Xu *et al.* [18]   (g) Xu *et al.* [17]
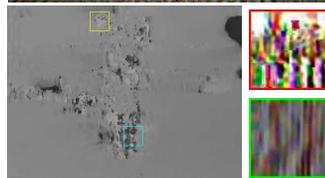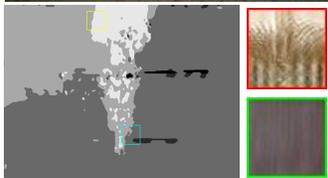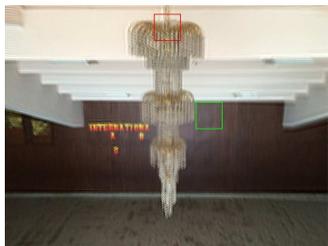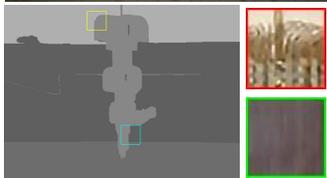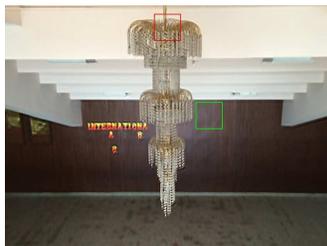
(e) Mathamkode *et al.* [1]   (f) Hu *et al.* [5]   (g) Nimisha *et al.* [11]

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
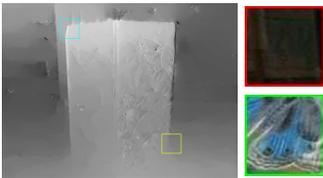
(b) Ours   (c) Mohan *et al*. [9]   (d) Pan *et al*. [12]
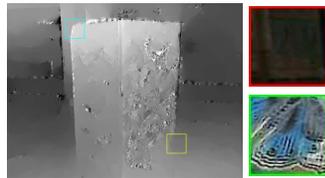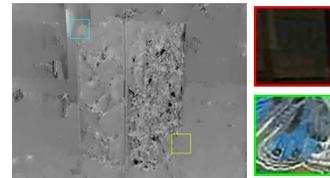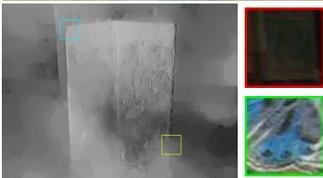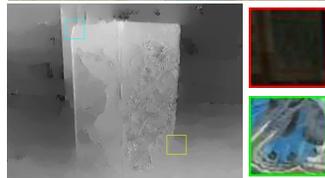
(e) Tao *et al*. [15]   (f) Xu *et al*. [18]   (g) Xu *et al*. [17]
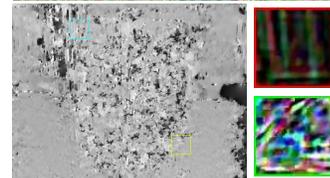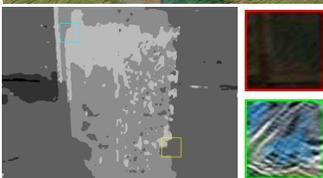
(e) Mathamkode *et al*. [1]   (f) Hu *et al*. [5]   (g) Nimisha *et al*. [11]

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]
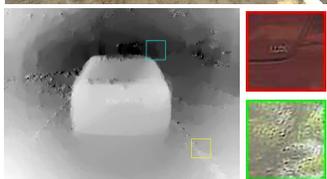
(e) Tao *et al.* [15]   (f) Xu *et al.* [18]   (g) Xu *et al.* [17]
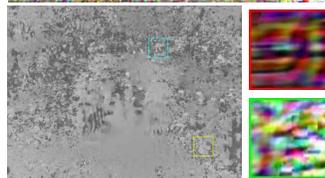
(e) Mathamkode *et al.* [1]   (f) Hu *et al.* [5]   (g) Nimisha *et al.* [11]

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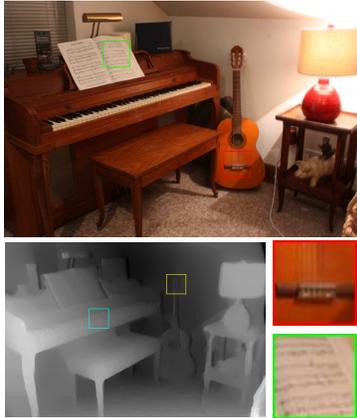
(b) Ours
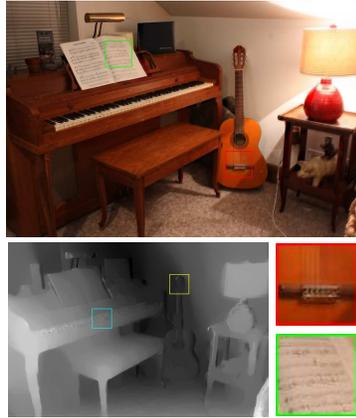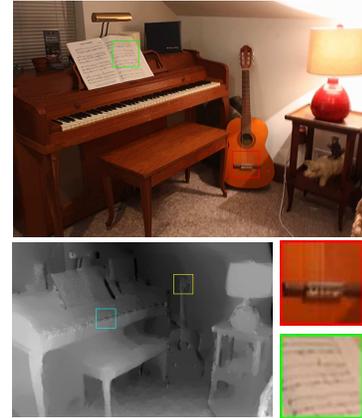
(c) Mohan *et al.* [9]

(d) Pan *et al.* [12]
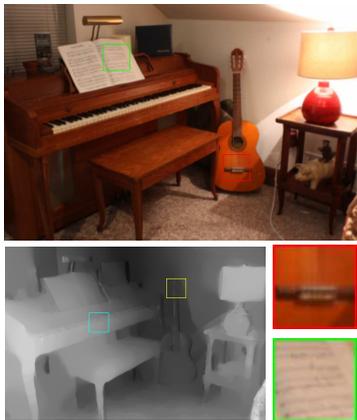
(e) Tao *et al.* [15]
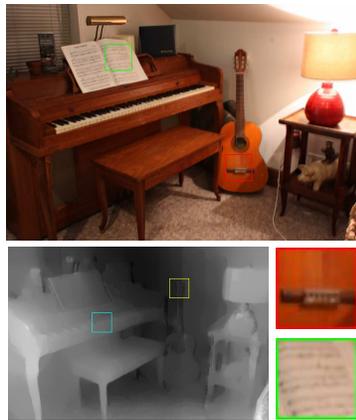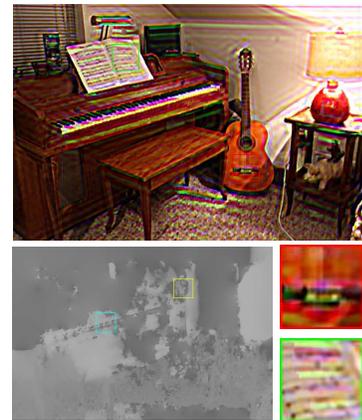
(f) Xu *et al.* [18]

(g) Xu *et al.* [17]
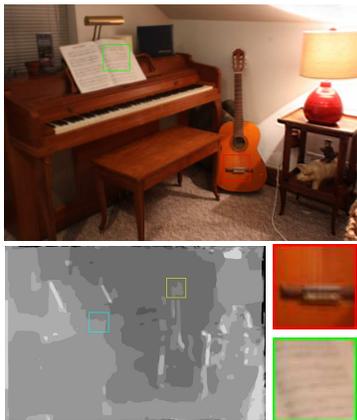
(e) Mathamkode *et al.* [1]
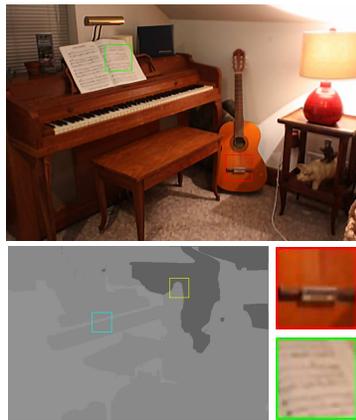
(f) Hu *et al.* [5]
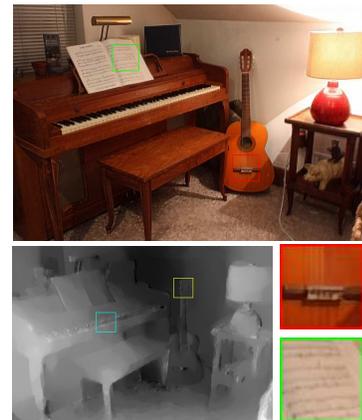
(g) Nimisha *et al.* [11]

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.
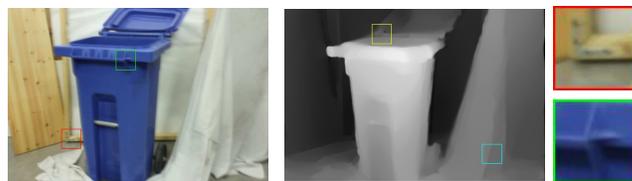
(a) Blurry image

(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 aritificial 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.
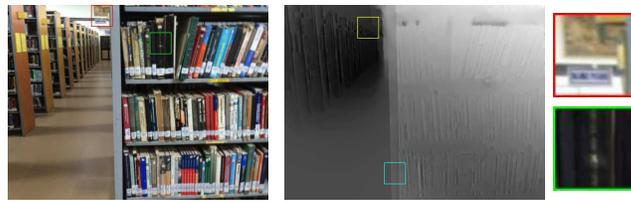
(a) Blurry image
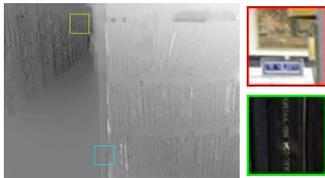
(b) Ours | (c) Mohan *et al.* [9] | (d) Pan *et al.* [12]
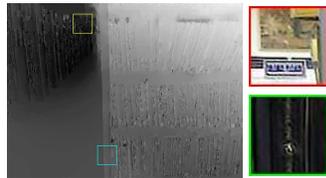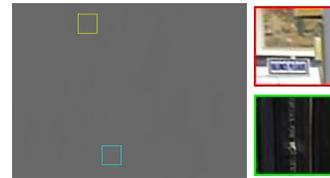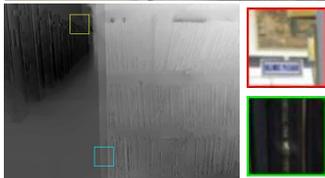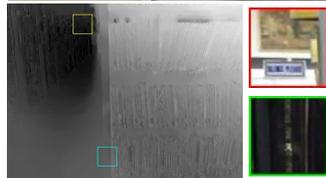
(e) Tao *et al.* [15] | (f) Xu *et al.* [18] | (g) Xu *et al.* [17]
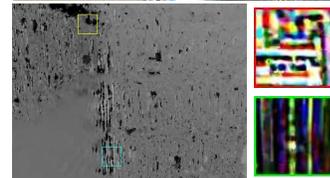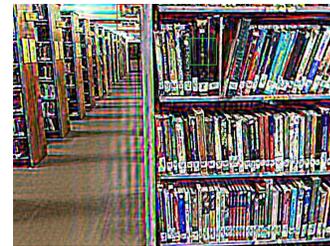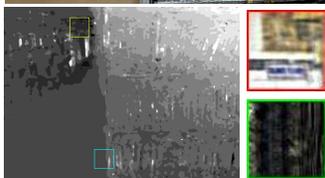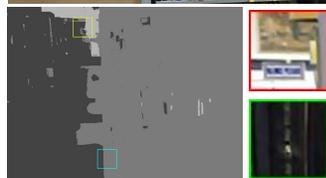
(e) Mathamkode *et al.* [1] | (f) Hu *et al.* [5] | (g) Nimisha *et al.* [11]

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.

(a) Blurry image   (b) Ours

(a) Blurry image   (b) Ours

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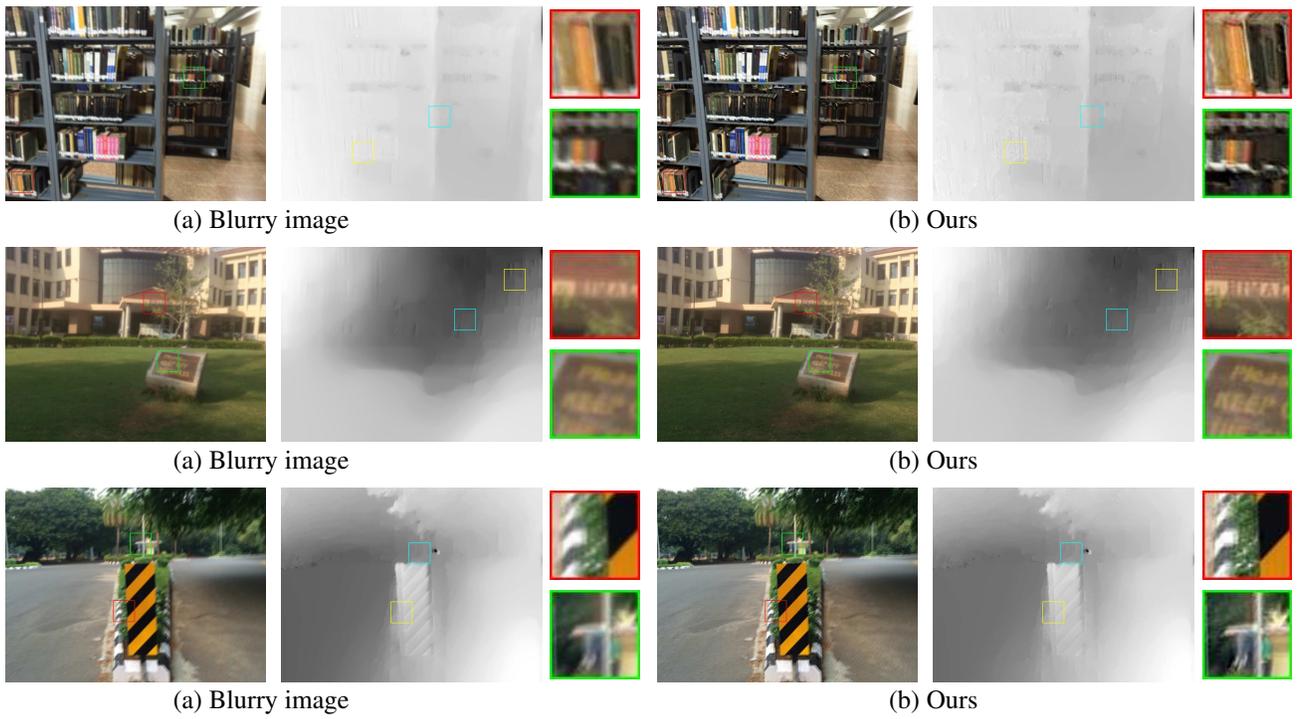
Figure S13. More real-world examples: (a) Blurry Image and (b) Our result.