

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 stereoscopic [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

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b, \quad (1)$$

where R is the corresponding rotational matrix [48], \mathbf{l}_b is the baseline vector ($\mathbf{l}_b = 0$ 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

$$\begin{aligned} \mathbf{I}_B^w &= \frac{1}{t_e^w} \int_{t \in t_e^w} P^w(R_t(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c) dt, \\ \mathbf{I}_B^n &= \frac{1}{t_e^n} \int_{t \in t_e^n} P^n(R_t(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b) dt, \end{aligned} \quad (2)$$

where the wide-angle camera is considered as reference (without loss of generality). In practice, the COR (\mathbf{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 $P(R_t(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b)$ we mean a transformation mapping $T_{(R_t, \mathbf{l}_c, \mathbf{l}_b)} : \mathbf{X} \rightarrow R_t(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b, \forall \mathbf{X}$ 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

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

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

$$\mathbf{x}' = \lambda \left(K^n R(K^n)^{-1} \mathbf{x} + \underbrace{\frac{1}{Z} K^n (I - R) \mathbf{l}_c}_{\text{center-of-rotation}} + \underbrace{\frac{1}{Z} K^n (I - R) \mathbf{l}_b}_{\text{baseline}} \right), \quad (4)$$

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 $\mathbf{l}_b = 0$, 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 $L = L^n + L^w$:

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

where $\|\mathbf{A}^k \mathbf{w}^k - \mathbf{I}_B^k\|_2^2 = \|\mathbf{M}^k \mathbf{I}_C^k - \mathbf{I}_B^k\|_2^2$.

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 i th 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 i th column contains the PSF corresponding to the i th 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*.

Proof: 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

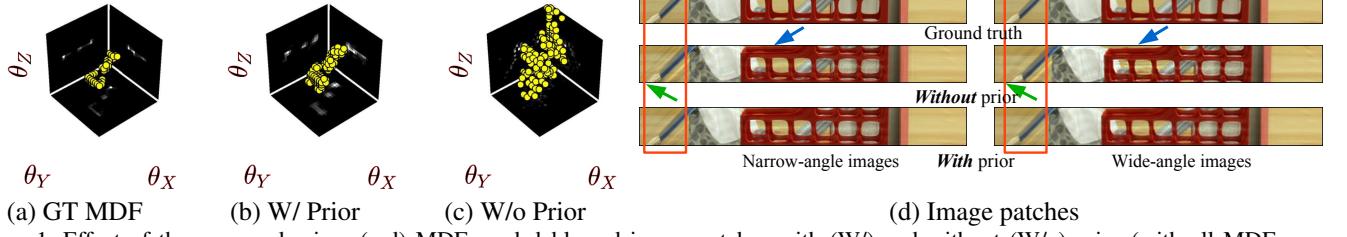


Figure 1. Effect of the proposed prior: (a-d) MDFs and deblurred image patches with (W/) and without (W/o) prior (with all MDFs centroid-aligned with the ground truth (GT) w^n to align left-images). MDF estimate of the prior-less case has a random offset (Fig. (c)) and the corresponding deblurred image 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

$$\begin{aligned} \mathbf{I}_B^n &= \sum_p w^n(p) P^n(R_p R_n^{-1} \underbrace{R_n(\mathbf{X} - \mathbf{l}_c)}_{\text{true}} + \mathbf{l}_c + \mathbf{l}_b), \\ &= \sum_p w^n(p) P^n(R_p R_n^{-1} \underbrace{(R_n(\mathbf{X} - \mathbf{l}_c) + \mathbf{l}_c - \mathbf{l}_c) + \mathbf{l}_c + \mathbf{l}_b}_{\text{apparent}}), \end{aligned} \quad (6)$$

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 $\{A, B, C\}$

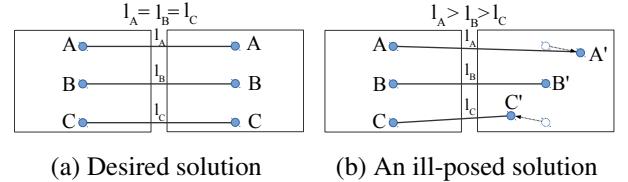


Figure 2. $\{A, B, C\}$ in Fig. (a) correspond to scene-features at the same depth (*i.e.*, *identical* disparities). Fig. (b) considers an inplane rotational ambiguity, wherein $\{A, B, C\}$ translates to $\{A', B', C'\}$ which clearly leads to *inconsistent* disparities.

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 $\tilde{l}_c = \arg \min_{l_c} (L_{l_c}^w + L_{l_c}^n)$:

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

where $k \in \{w, n\}$, $g(\cdot)$ produces the first and second-order gradients, and the symbol ‘~’ 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

$$\begin{aligned} \arg \min_{\mathbf{w}^n} \|\tilde{\mathbf{A}}^n \mathbf{w}^n - \mathbf{I}_B^n\|_2^2 + \alpha \|\mathbf{w}^n - \tilde{\mathbf{w}}^w\|_2^2 : \|\mathbf{w}^n\|_1 \leq \lambda_1'^n, \\ \arg \min_{\mathbf{I}_C^n} \|\tilde{\mathbf{M}}^n \mathbf{I}_C^n - \mathbf{I}_B^n\|_2^2 + \lambda_2^n \|\nabla \mathbf{I}_C^n\|_1, \end{aligned} \quad (8)$$

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.

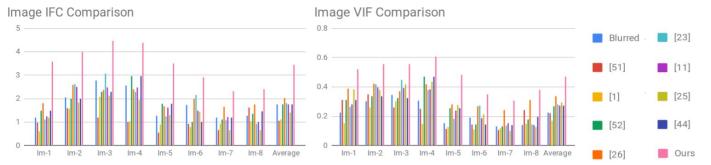


Figure 4. Quantitative evaluations using subjective measures (IFC, VIF). Our method performs deblurring with the best aesthetics.

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-

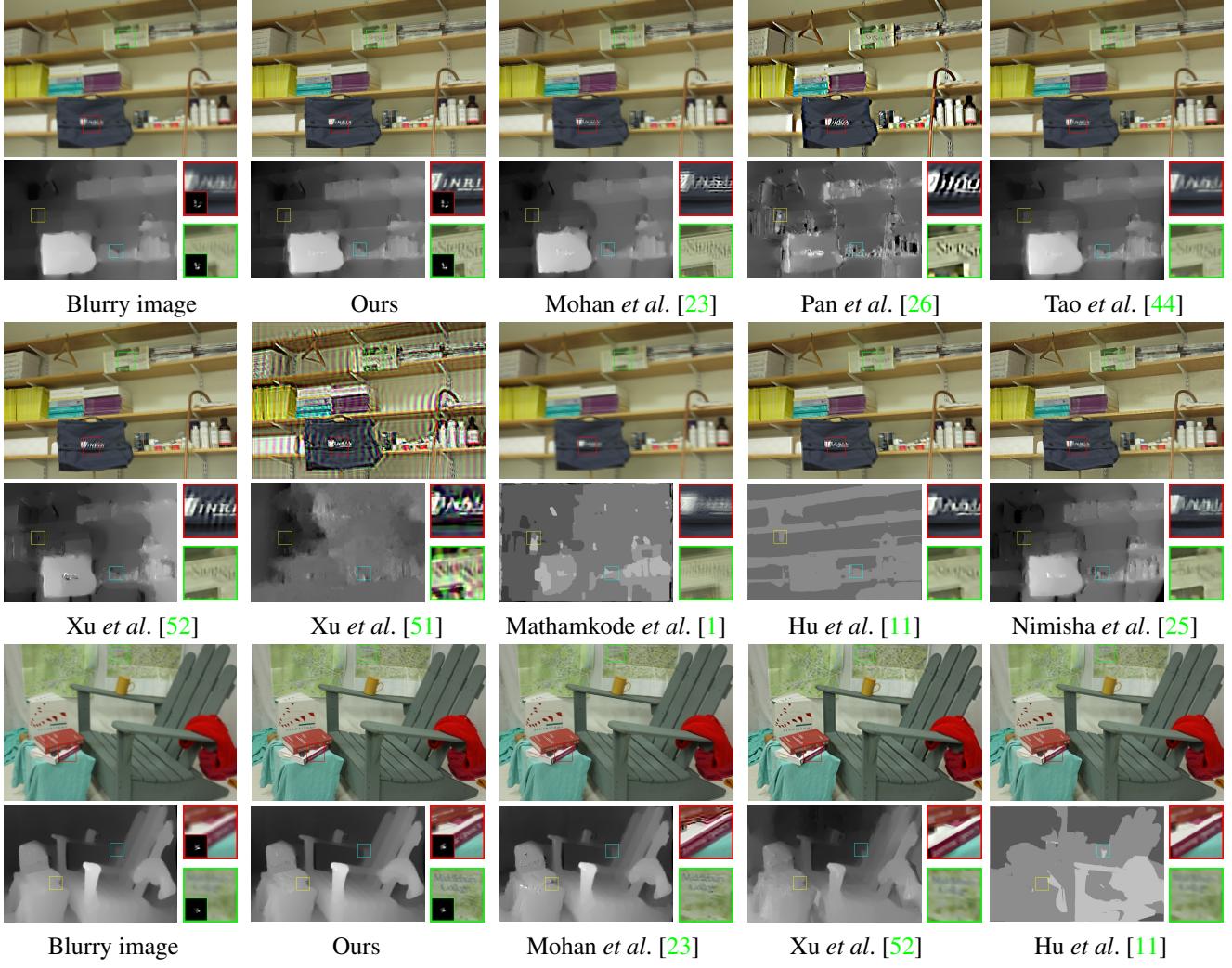


Figure 5. Synthetic experiments: The method of [51, 11, 1] exhibits severe ringing artifacts and inaccurate depth estimates. The results of [26, 52] amply underline the shortcomings of normal camera models. As compared to deep learning [44, 25] and light field BMD [23], our method retrieves distinct 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

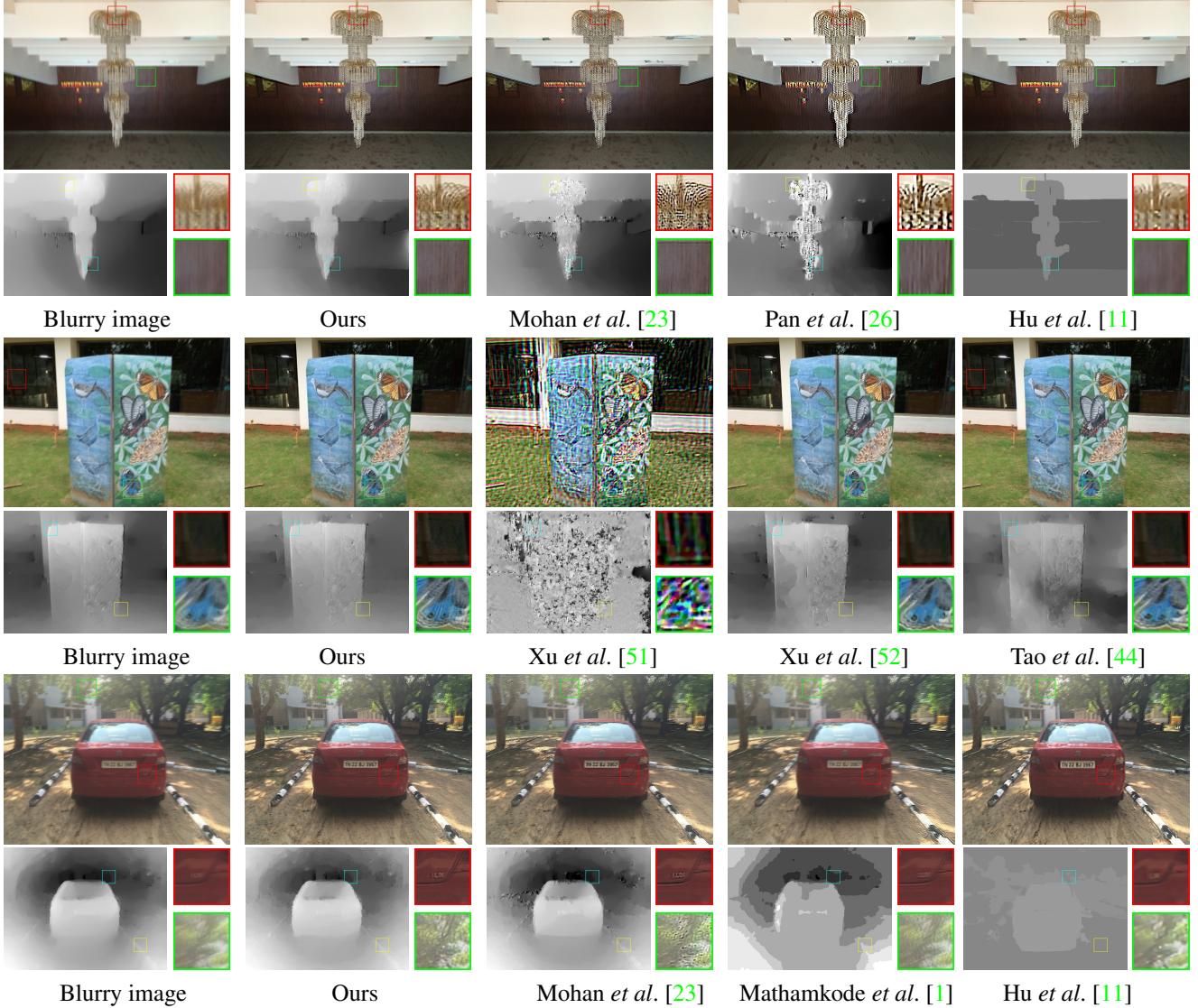


Figure 6. Real experiments: (first row - indoor scene, second and third row - outdoor scenes). Our method is able to recover finer features at different depth ranges as compared to the competing methods, and is able to faithfully preserve the depth information.

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 pro-

cess 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.

Acknowledgement: The first author gratefully acknowledges travel grant from Google, & PhD scholarship from MHRD India.

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 mea-
- surement 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 Sroubek, 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