

Divide and Conquer for Full-Resolution Light Field Deblurring

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

Abstract

The increasing popularity of computational light field (LF) cameras has necessitated the need for tackling motion blur which is a ubiquitous phenomenon in hand-held photography. The state-of-the-art method for blind deblurring of LFs of general 3D scenes is limited to handling only downsampled LF, both in spatial and angular resolution. This is due to the computational overhead involved in processing data-hungry full-resolution 4D LF altogether. Moreover, the method warrants high-end GPUs for optimization and is ineffective for wide-angle settings and irregular camera motion. In this paper, we introduce a new blind motion deblurring strategy for LFs which alleviates these limitations significantly. Our model achieves this by isolating 4D LF motion blur across the 2D subaperture images, thus paving the way for independent deblurring of these subaperture images. Furthermore, our model accommodates common camera motion parameterization across the subaperture images. Consequently, blind deblurring of any single subaperture image elegantly paves the way for cost-effective non-blind deblurring of the other subaperture images. Our approach is CPU-efficient computationally and can effectively deblur full-resolution LFs.

1. Introduction

Handheld light field cameras (LFCs) are being used in a variety of tasks including wide-angle and hyperspectral imaging, shape recovery, segmentation, etc. [4, 25, 14, 22]. Refer to [24] for a recent survey on light field (LF) related works. The increase in popularity of LFCs can be attributed to their attractive features over conventional cameras (CCs), including post-capture refocusing, f-stopping, depth sensing [22, 1, 16], etc. LFCs achieve this by capturing multiple (subaperture) images instead of a single CC image by segregating the light reaching the CC-sensor into multiple angular components; and synthesize these images post-capture to form an image of desired CC setting [16, 1]. However, there is a downside too. It is well-known that motion blur is a common artifact in hand-held photography with CCs

[17, 26, 12]. This nuisance effect become exacerbated in LFCs. This is because the light-segregation principle in LFCs reduces the amount of photons that make up individual subaperture images, thereby necessitating higher exposure times relative to CC (under the same setting). This escalates the risk of motion blur in LFC. Moreover, a 4D LF comprising of 2D spatial and 2D angular resolutions can be interpreted as several CC images stacked together. Thus the numerical optimization involved in LF deblurring must deal with very large-sized data as compared to that of CC. This poses additional computational challenges [24, 20].

In this work, we address the problem of LF blind motion deblurring (LF-BMD). The problem of BMD deals with estimating both the clean image and underlying camera motion from a *single* motion blurred observation. BMD in CCs is a well-studied topic replete with efficient methodologies. State-of-the-art CC-BMD methods [17, 26, 21] are based on the motion density function (MDF) [5] which allows both narrow- and wide-angle systems as well as nonparametric camera motion, have a homography-based filter flow framework for computational efficiency [8], and employ a scale-space approach to accommodate large blurs. Kohler *et al.* [11] have shown (for CCs) that general camera motion comprising of 3D translations and 3D rotations can be well approximated by full rotations, or inplane translations and inplane rotation. Inplane rotation common to both the approximations are necessary to capture wide angle settings [15, 21]. Whyte *et al.* [23], Xu *et al.* [26], and Pan *et al.* [17] follow full-rotations approximation and employ EFF to yield high quality results in CC-BMD.

In contrast, LF-BMD is an emerging research area and there exists very few works. Jin *et al.* [10] proposed the first LF-BMD approach, but restrict the scene to be fronto-parallel and bilayer, and limit the camera motion to *only* inplane translations. A recent LF-BMD work by Srinivasan *et al.* [20] eliminates the planar scene assumption and even includes full 3D translations. However, the ego-motion is constrained to be parametric. This *reduces* its effectiveness under irregular ego-motions, which is common when imaging from moving vehicles, robotic platforms, etc. Moreover, since the translational pose cannot model inplane ro-

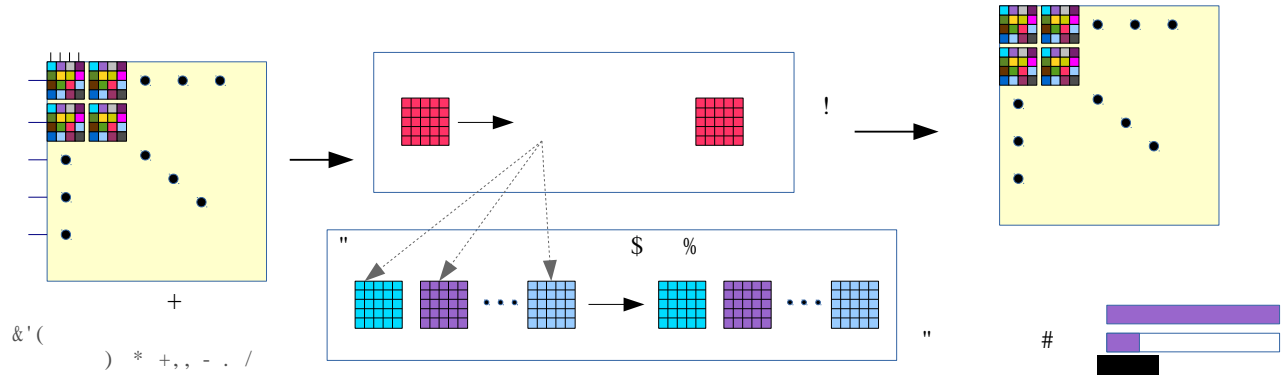


Figure 1. Outline of our proposed method. Our LF-BMD enables decomposing 4D LF deblurring problem into a set of *independent* 2D deblurring sub-problems, in which a blind deblurring of a *single* subaperture-image enables *low-cost* non-blind deblurring of individual subaperture images in parallel. Since all our sub-problems are 2D (akin to CC-case) and thus cost-effective (as it allows efficient filter flow or EFF [8] and is CPU-sufficient), our method is able to handle full-resolution LFs, with significantly less computational cost.

tation, both [10] and [20] are *ineffective* for wide angle systems. Dansereau *et al.* [2] introduce a hardware-assisted deblurring approach using a robotic arm to mount the camera and estimate camera motion (hence non-blind). To the best of our knowledge, [20] is the only work which deals with blind motion deblurring for general 3D scenes. Here, LFC is modeled as an array of pinhole cameras by discarding the effect of LFC-lens, and the motion blurred 4D LF is treated as a composition of shifted and sheared versions of a clean 4D LF. Deblurring with this model proceeds by optimizing for a clean 4D LF at ‘one go’, using a 4D prior [20].

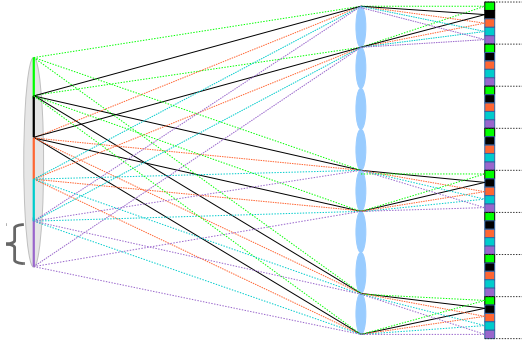
However, [20] has some major drawbacks. First, optimization of 4D LF *in toto* brings up new challenges. The computational requirement involved in this optimization restricts [20] to handle *only* downsampled LFs – both in spatial and angular resolutions (*e.g.*, a Lytro Illum LF file decoded using [3] has 197 subaperture images of size 433×625 , whereas [20] requires downsampling it to 64 images of size 200×200 for computational feasibility). As LF post-capture rendering involves composition of multiple subaperture images (or angular components), angular downsampling can adversely affect rendering performance. On the other hand, spatial downsampling restricts LF-rendering software to produce *only* low-resolution RGB images. Furthermore, optimizing high-dimensional data elevates the computational complexity and the method warrants GPU-based processing. Also, such a high-dimensional optimization can distort the interrelations among subaperture images due to convergence issues, which is an important factor for consistent post-capture rendering of LFs [16].

In this paper, we introduce an MDF-based LF motion blur model which allows for decomposition of LF-BMD into low-dimensional subproblems. This admits an efficient filter flow framework [8] to remove the computational bot-

tlenecks and several other limitations of the state-of-the-art [20]. Specifically, our model *isolates* blur formation in *individual* subaperture images (unlike [20]), and imparts a *dependency* among *all* subaperture images through a *common* MDF. Our formulation performs LF-BMD in two efficient steps, as illustrated in Fig. 1. First, we estimate the common MDF from the center-subaperture image using BMD (akin to CC-BMD). Second, by invoking the blur-isolation and commonality of MDF properties inherent to LFCs, we perform *independent* (or parallelizable) non-blind deblurring of individual subaperture images using the estimated MDF while simultaneously accounting for the lens-effect and parallax arising from separation of subapertures from the lens-center. Since each of these subproblems is low-dimensional, our method overcomes the drawbacks associated with the high-dimensional optimization, and thus can deblur LFs at full-resolution [20]. In addition, unlike [20], our LF-MDF model captures the effect of camera lens, can cater to both wide- and narrow-angle camera settings, and can handle irregular camera motions.

Our main contributions are summarized below.

- By harnessing the physics behind LF, we decompose 4D LF-BMD to 2D subproblems, which enables the first ever attempt of full-resolution LF-BMD.
- Our work bridges the gap between the well-studied CC-BMD and emerging LFC-BMD, and facilitates mapping of analogous techniques (such as MDF formulation, efficient filter flow framework, and scale-space strategy) developed for the former to the later.
- Our work dispenses with some important limitations impeding the state-of-the-art [20], such as high computational cost, GPU requirement, and ineffectiveness in handling wide-angle systems & irregular ego-motions.



(a) Formation of 4D light field

(b) Formation of motion blur in a subaperture image

Figure 2. LF motion blur model: (a) Formation of 4D LF. (b) Interpreting camera motion as relative world motion, each motion blurred 2D subaperture image is obtained as a combination of the projections of moving world (parametrized by an MDF) through the *respective* subaperture onto a virtual sensor or microlens array. Also, all subapertures experience the *same* world motion (or share a *common* MDF).

2. MDF for Light Field

In this section, we discuss the working of LFC and limitations of the LF motion blur model of [20]. We then proceed to conceptualize (akin to conventional cameras) an MDF based interpretation for motion blur in LFs, that seeks to mitigate the drawbacks of [20].

A CC with a full-aperture setting spatially resolves light onto a 2D sensor array. An LFC further segregates the light in each CC sensor-element in accordance with the angular bin of the lens-aperture (or subaperture) through which the light arrives. This is achieved by means of a microlens placed in the position of each CC sensor element, which resolves each spatial component into multiple angular components, as shown in Fig. 2(a). The 2D spatially as well as 2D angularly resolved light is stored in a high-resolution LF-sensor (behind the microlens array) to form a 4D LF.

The LF motion blur model of [20] discards the effect of LF-lens and approximates LFC as an array of pinhole cameras positioned at the subapertures. The camera motion is interpreted as an associated movement of these pinhole cameras. In effect, a motion blurred LF is modeled as a composition of shifted and sheared versions of the clean 4D LF. Using flatland analysis (*i.e.*, considering a single angular dimension u and a single spatial dimension x), a motion blurred 4D LF (L_b) can be represented as

$$L_b(x, u) = \int_t L_c(x, u + p_x(t) - xp_z(t)) dt, \quad (1)$$

where L_c is the clean 4D LF and $\{p_x(t), p_z(t)\}$ is the camera motion path during the exposure time. As only the angular term u is varying in Eq. (1), a motion blurred subaperture image can be interpreted as a composition of multiple clean subaperture images (where the amount depends on camera motion). In the above equation, considering a *single*

blurred 2D subaperture image as the observation amounts to solving for *multiple* clean 2D images as unknowns – a heavily ill-posed problem. Instead, [20] considers the entire blurred 4D LF as observation and solves for a clean 4D LF as unknown. This reduces the ill-posedness but incurs high-dimensional optimization issues, as discussed in Sec. 1.

We now briefly digress to discuss the MDF model employed in the state-of-the-art CC-BMD [17, 26, 23]. In this model, the CC is approximated as a pinhole at lens-center, and camera motion is interpreted as stationary camera but with relative world motion (see Fig. 3). Considering full-rotations approximation and a single pose change, the relative change in world coordinate is given as

$$X = RX, \quad (2)$$

where R is rotation matrix, and $X = [X, Y, Z]^T$ and $X = [X, Y, Z]^T$ are the 3D world coordinates with respect to initial and final camera positions, respectively. There exists a homography mapping h that relates the corresponding displacement in image coordinates, *i.e.*,

$$x = h(K, R, x), \quad (3)$$

where K is the camera matrix, and x and x are 2D image coordinates corresponding to the initial and final camera positions, respectively. Resultantly, the final image I can be related to the initial image I as

$$I = I(K, R), \quad (4)$$

where $I(K, R)$ performs warping of image I in accordance with Eq. (3). Thus a general motion blurred image B (*i.e.*, camera that experiences multiple pose-changes over its exposure time) can be expressed as

$$B = \cdot I(K, R), \quad (5)$$

where \mathbf{R} spans the plausible camera pose-space and ρ is the motion density function (MDF) which gives the fraction of exposure time the camera stayed in the pose \mathbf{R} (Fig. 3-right). Note that the MDF completely characterizes the camera-shake. State-of-the-art CC-BMD methods proceed by alternating minimization of \mathbf{B} and ρ in a scale-space manner to accommodate large blurs (*i.e.*, MDF estimation starts with a downsampled blurred image where the motion blur is less prominent, and proceeds to finer scale MDF-estimation using the previous estimate).

Along similar lines, it is possible to conceptualize camera shake in LFC to be uniquely characterised by an MDF, but having one-to-many mapping from world to LF-sensor due to LF-capture mechanism. To develop an analogous MDF framework for LF-BMD, we introduce an alternative interpretation of LF motion blur formation. The design of LFC-lens system is such that light from different subapertures do *not* interfere with each other [16]. As shown in Fig. 2(b)-right, a clean subaperture image is equivalent to a 2D image formed in a CC with full-aperture setting and with *only* the respective subaperture open, the sensor being placed at the position of the microlens-array.

Akin to CC, we too interpret motion to be stationary camera and a world moving (see Fig. 2(b)). Each motion blurred subaperture image is thus equivalent to an image formed in a full-aperture CC with *only* the respective subaperture open. Interestingly, note that all subapertures are subjected to same world-motion (*i.e.*, parameterized by a *single* MDF). Intuitively, *each motion blurred subaperture image is formed by a linear combination of the projections of the moving world through the respective subaperture onto a virtual sensor formed by the microlens array*. Thus, a blurred subaperture image \mathbf{B}_i can be alternatively expressed as

$$\mathbf{B}_i = \int \rho(\mathbf{R}) \cdot \mathbf{I}_i(\mathbf{K}_i, \mathbf{R}, \mathbf{X}) d\mathbf{R} \quad 1 \leq i \leq N, \quad (6)$$

where \mathbf{I}_i is the clean version of the i th subaperture image. Note that the camera matrix \mathbf{K}_i and parameter set \mathbf{X}_i vary with individual subapertures so as to capture one-to-many world-mapping. Their exact forms will be discussed in a later section. Our model isolates the blur in individual subaperture images, *i.e.*, a *single* 2D blurred subaperture image as observation amounts to solving for the *corresponding* clean subaperture image as unknown (unlike [20]). Also, a *single* MDF ρ is shared by all subaperture images.

The MDF-based LF motion blur model in Eq. (6) provides three distinct advantages. First, because it *isolates* motion blur in individual subaperture images, we can estimate the common MDF from a *single* subaperture image – a low-dimensional optimization (akin to CC-BMD). Second, since all subaperture images share a *common* MDF, we can use the estimated MDF to perform non-blind de-

Figure 3. Motion Density Function (MDF): Change in camera orientation from A to B is equivalent to the relative change in world coordinate system (CS) from \mathbf{X} to \mathbf{X}' . Thus, MDF, which gives the fraction of time the world CS stayed in different poses during the exposure time, *completely* characterizes the camera motion.

blurring of all the other subaperture images. As non-blind deblurring of individual subapertures can be done *independently*, this step is amenable to parallelization. Note that non-blind deblurring methods (which optimize for a clean 2D image *only* once) are quite cost-effective as compared to blind methods (which clumsily optimize for MDF and clean 2D image alternately over iterations). These factors drastically reduce the computational cost for LF-BMD and allow full-resolution LF-BMD. Third, since MDF captures both regular and irregular ego-motion, our method can handle unconstrained ego-motion; and consideration of full rotational camera motion accommodates both narrow- and wide-angle systems, unlike [20, 10].

3. MDF-based LF Motion Blur Model

In this section, we formulate our MDF-based LF motion blur model. This requires world-to-sensor mapping in each subaperture, so as to derive individual LF homographies.

3.1. World-to-Sensor Mapping in a Subaperture

Conventional cameras with a small-aperture setting can be well-approximated by a pinhole centered at the aperture’s center. This approximation is widely used in many practical applications (including CC-BMD) [6, 17, 26]. In LFCs, the characteristics of light refraction over different subapertures vary in accordance with their positions due to the effect of large-aperture lens. This effect cannot be captured with a pinhole array (as the main lens is not involved); *e.g.*, a beam of parallel rays through LFC-lens *converge* at the focal point, but will pass *parallel* through a pinhole camera array. To account for this effect, we approximate subapertures as pinholes over subaperture-centers, and yet configured to obey the refractions incurred at that portion.

Fig. 4(a) shows a flatland ray tracing model for a subaperture positioned above the optical-center and a world point with positive X coordinate. Following the thin-lens equation with focal length f , a light-ray from a world point $\{\mathbf{X}_s, \mathbf{Y}_s, \mathbf{Z}_s\}$ through the subaperture has to pass through the point of intersection of the principal ray (*i.e.*, a ray

We also throw light on the possibility of individually deblurring subaperture images using CC-BMD. Assuming $K_{k_{xy}} = K$ and neglecting k_{xy} necessitates different MDFs for capturing the one-to-many mapping of LF. This distorts their mutual consistencies (*e.g.*, see epipolar image or EPI of Figs. 7(d-e)), mainly due to shift-ambiguity of latent image-MDF pair [15], and relative estimation-error of different MDFs. This adversely affects the refocusing and f-stopping functionality of LFs [16]. Furthermore, since blind deblurring is very costlier than non-blind, the computational cost climbs steeply with spatial and angular resolutions.

4. Optimization of LF-BMD

In this section, we discuss our LF-BMD approach that comprises of two steps: First, estimate the *common* MDF from a *single* subaperture image using BMD, and second, employ the estimated MDF to perform low-cost non-blind deblurring of *remaining* subaperture images (in parallel).

4.1. LF-MDF Estimation

The homography mapping of the center-subaperture image (*i.e.*, $k_x = k_y = 0$ in Eq. (14)) is equivalent to that of a CC-pinhole model [6], *i.e.*,

$$\mathbf{x} = \frac{Z}{Z_0} \mathbf{K} \mathbf{R} \mathbf{K}^{-1} \mathbf{x}_0 \quad \mathbf{K}_{k_{xy}} = \mathbf{K} \text{ and } \mathbf{b}_{k_{xy}} = 0, \quad (16)$$

where scalar $\frac{Z}{Z_0}$ ($= Z/Z_0$) normalizes the third coordinate of \mathbf{x} to unity (Eq. (11)). Note that even though depths Z and Z_0 are present in Eq. (16), it is *not* required for homography mapping (and thus for MDF estimation) since it translates to a normalization of the third coordinate of \mathbf{x} to unity (see structure of \mathbf{x} in Eq. (11)) through [6, 17, 26, 23]. Thus, any state-of-the-art CC-BMD method can be employed to find the LF-MDF using the center-subaperture image.

4.2. EFF for Non-Blind Deblurring of LFs

Since a common MDF is shared among all subaperture images, we utilize the estimated MDF to perform non-blind deblurring of individual subaperture images. For a non-centered subaperture, camera matrix $K_{k_{xy}}$ varies with subaperture positions and the additive term of Eq. (14) is nonzero (which makes it different from the CC-pinhole case). Eventhough $1/Z$ in Eq. (14) can be obtained by normalization (as in CC-case), the depth information Z is required for homography mapping to capture parallax and lens effect. A direct approach for non-blind deblurring involves constructing a large matrix $M_{k_{xy}}$ using the MDF formulation of Eq. (15) for each subaperture k_{xy} , to solve the optimization problem

$$\hat{I}_{k_{xy}} = \min_{I_{k_{xy}}} M_{k_{xy}} I_{k_{xy}} - B_{k_{xy}} \frac{1}{2} + \text{prior}(I_{k_{xy}}), \quad (17)$$

where ‘prior’ is an image regularizer, such as total variation (TV), sparsity in image gradient [26], dark channel [26],

etc. As a full-resolution LF is composed of numerous subaperture images, construction of $M_{k_{xy}}$ and optimization of individual subaperture images with priors are computationally expensive. To this end, we elegantly extend the efficient filter flow (EFF) employed in CCs [8] to LFCs.

The EFF approximates space-variant blur in an image to be locally space-invariant over small image patches. Using this approximation, we can simplify the blurring process in a subaperture image as

$$B_{k_{xy}} = \sum_{i=1}^R C_i^\dagger \cdot h_{k_{xy}}^i (C_i \cdot I_{k_{xy}}), \quad (18)$$

where i iterates over R overlapping patches in clean subaperture image $I_{k_{xy}}$, $C_i \cdot I$ is a linear operation which extracts the i th patch from I , $(h \cdot C_i \cdot I)$ performs a convolution with kernel h on i th patch, and C_i^\dagger inserts the patch back to its original position with a Barlett windowing operation. The convolution kernel $h_{k_{xy}}^i$ corresponding to the i th patch center can be derived using Eq. (15) as

$$h_{k_{xy}}^i = C_i \cdot \frac{1}{R} \cdot (K_{k_{xy}}, R, k_{xy}), \quad (19)$$

where i is an image of the same size as that of the subaperture image with only an impulse located at the i th patch center. EFF requires MDF-based motion blur model to be calculated *only* at patch centers and eliminates the need for building large matrices for optimization, as in Eq. (17). The EFF allows for an efficient patch-based deblurring:

$$\hat{I}_{k_{xy}} = \sum_{i=1}^R C_i^\dagger \cdot \text{deconv} (h_{k_{xy}}^i, (C_i \cdot B_{k_{xy}})), \quad (20)$$

where ‘deconv’ indicates non-blind deconvolution which is computationally efficient as compared to blind deblurring.

5. Analysis and Discussions

In this section, we elaborate on the validity of rotation-only approximation in LF-BMD, and depth estimation.

Rotation-only approximation: To reduce the number of unknowns in MDF, CC-BMD methods typically approximate full 6D motion to 3D [17, 23]. As the effect of translation is negligible for normal camera-shakes [23], 3D rotation-only approximation is commonly employed in CCs. We show that this is valid for LFCs too. Considering the worst-case (WC) *plausible* camera translation as $T = [|r|, |r|, |r|]$, [23] shows for CC that the corresponding WC pixel translation $t = [t_x, t_y, t_z] = K_0 T / Z$ can be *ignored* ($K_0 = \text{diag}(u, u, 1)$). We claim that pixel translations in our LFC-model is *equivalent* to that of CC for inplane translations, and *approximates* t for 3D T . Considering $X = RX + T$ in Eq. (13), T in SA i (using Eq. (14))

Figure 5. Evaluation of depth estimation cues.

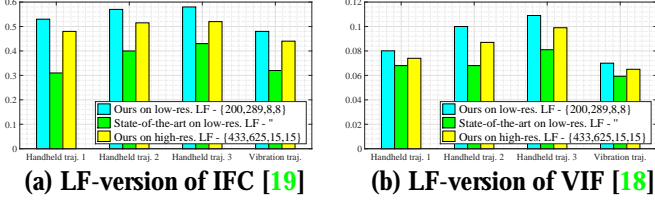


Figure 6. Quantitative evaluation for wide-angle setting ($f = 29$ mm) using real hand-held trajectories (from [11]) and irregular camera motion using vibration trajectory (from [7]).

amounts to $\hat{\mathbf{t}} = \mathbf{K}_i \mathbf{T} / \mathbf{Z} = (\mathbf{K}_0 + \mathbf{M}_i) \mathbf{T} / \mathbf{Z}$, where $\mathbf{M}_i = [0, 0, k_x(f - u)/f; 0, 0, k_y(f - u)/f; 0, 0, 0]$. For $\mathbf{T}_z = 0$, $\mathbf{M}_i \mathbf{T} / \mathbf{Z} = 0$ (equal effect in CC & LFC). Note that WC displacement happens for highest k_x or k_y ($= f/4$). For 3D \mathbf{T} , LFC $\hat{\mathbf{t}} = [t_x, t_y, t_z]$, where $t_x = 1 + |(f - u)/4u|$ or $1 + |f|/(4|Z|)$ (using Eq. (7)). As $|Z|$ is in the order of m and $|f|$ in mm , $1 < 2$ (i.e. $\hat{\mathbf{t}} \approx \mathbf{t}$). Hence proved.

Depth Estimation: To estimate the convolution kernels for different SAIs (for EFF), our algorithm requires depth information (in Eq. (19)). We use [22] to estimate depth for each patch by picking the most-confident depth estimate within that patch (without final depth refinement). Our consideration of uniform depth within a small image patch is analogous to the flatness and global smoothness priors commonly used for final depth-refinement [22, 9]. Depth estimation method in [22] is as follows. Refocusing LF translates to a skew in EPI, and the features of EPIs for a image point will be vertical (or horizontal depending on projection) when it is at focus [16]. [22] skews EPIs corresponding to different depths, and picks among them the depth which makes EPIs vertical. Motion blurred LFs also possess EPI characteristics required for depth estimation (compare EPIs in Figs. 7-8(a) and depth cues in Fig. 5).

6. Experimental Results

In this section, we provide quantitative and qualitative evaluations to highlight the computational gain of our approach and its ability to deal with full-resolution LFs with competitive performance. We also show that our method can deal with both wide-angle systems and irregular camera trajectories, unlike the state-of-the-art LF-BMD [20].

Datasets used: For real experiments on low-resolution LFs, we used the motion blurred LF dataset of [20]. Since there exist no full-resolution motion blur LF-datasets, we create one with LFs captured using **Lytro Illum** and decoded raw-LFs to **MATLAB** format full-resolution LFs using [4].

LF-resolution {x, y, u, v}	State-of-the-art [20] (GPU-based)	Ours (CPU-based)
{200, 200, 8, 8} (Dataset of [20])	2 hrs, 20 mins	8.21 mins (Gain 17.05×)
{200, 289, 8, 8} (Low-res. LF)	3 hrs, 17 mins	12.62 mins (Gain 15.61×)
{433, 625, 15, 15} (Full-res. LF)	Not feasible (Resource allocation error)	38 mins (Feasible)

Table 1. Time comparisons. *Over 90% of the time is used for *low-cost* 197 non-blind deblurring parallelized in 8 cores of a CPU. Using more cores or GPU further improves the speed significantly.

For quantitative evaluation, we synthesized motion blur on clean full-resolution LFs using real handheld trajectories from [11] with 29 mm focal-length (wide-angle setting) and 1/50 s exposure time. For irregular motion, we used real vibratory ego-motion trajectory from [7].

Comparison methods: We consider mainly the current state-of-the-art LF-BMD [20] for evaluation. To demonstrate the *ineffectiveness* of CC methods on LFs, we also use state-of-the-art CC-BMD methods [12] and [17] to perform *independent* deblurring on individual subaperture images. The codes for [20, 17] and [12] are downloaded from the authors' website, and used their default parameters for all methods. Additional analysis, comparisons, and implementation details are provided in our supplemental material.

Quantitative Evaluation: We introduce an LF-version of information fidelity criterion (IFC) [19] and visual information fidelity (VIF) [18], which are shown to be the best metrics for BMD evaluation in [13], by averaging these metric over subaperture images. As processing full-resolution LFs using [20] is not feasible, we use a downsampled version (by 0.5) of our dataset to perform comparisons with [20]. Using IFC/VIF, Figs. 6(a-b) compare with [20] for wide-angle scenario (using real trajectories of [11]) and irregular camera motion (using [7]). It is evident from Fig. 6 that our method performs better than [20] (performance degradation of [20] may be attributed to its inability to model these scenarios); ours can also deblur full-resolution LFs (unlike [20]). Table. 1 gives the timing comparisons with the state-of-the-art [20]. It is evident that, even though our method uses only CPU, we achieve a gain of atleast an order relative to the GPU-based [20]. Also, our method performs full-resolution LF-BMD within three-quarters of an hour, which can be further improved using more cores or GPUs.

Qualitative Evaluation: We qualitatively evaluate our deblurring performance on real LFs. Fig. 7 gives an example of low-resolution LF. Note that the EPIs of [20], [12] and [17] are not consistent with the input. Also, there exists ringing artifacts in Fig. 7(c) of [20] (especially in upper leaves). In contrast, our result in Fig. 7(b) reveals intricate

(a) Input (b) Ours (c) Srinivasan *et al.* [20] (d) Krishnan *et al.* [12] (e) Pan *et al.* [17]

Figure 7. Comparison using low-resolution LF ($\{200, 200, 8, 8\}$) from dataset of [20]. (a) Input, (b) Ours, (c) State-of-the-art LF-BMD [20], (d) State-of-the-art CC-BMD [12] (e) State-of-the-art CC-BMD [17]. Note the inconsistencies in EPI w.r.t input for c (possibly due to convergence issues) and d-e (possibly due to *lack* of dependency among BMD of subaperture images). Also, notice the ringing artifacts in the upper leaves in c. In contrast, ours reveals more details (like veins of lower leaf), has negligible ringing artifacts, and EPI is consistent.

(a) Input (b) Ours (c) Srinivasan *et al.* [20] (d) Pan *et al.* [17]

Figure 8. Comparisons using full-resolution LF ($\{433, 625, 15, 15\}$) of **Lytro Illum**. Top-row shows a well-lit case and bottom row shows a low-light scenario. (a) Input, (b) Ours, (c) State-of-the-art LF-BMD [20] and (d) State-of-the-art CC-BMD [17]. [20] can *only* deblur downsampled LF due to computational constraints. Ours produce a superior full-resolution LF with consistent EPIs in all cases.

details (see the veins in lower leaf), has negligible ringing artifacts and produces consistent EPIs. Since this example is captured using low-light setting, noise is present in the input as well as our output. A detailed analysis to tackle noise in LF-BMD is presented in our supplementary. Fig. 8 shows comparisons with real full-resolution LFs, where the top and bottom rows depict well-lit and low-lit scenarios, respectively. The LF-BMD of [20] processes *only* a downsampled LF (both spatially and angularly) due to computational constraints. In contrast, the proposed method gives superior results in full-resolution and with consistent EPIs.

7. Conclusions

We introduced a novel interpretation of motion blur in 4D LF as *independent* blurring of multiple 2D images, yet

all sharing a *common* motion parametrization. This paved the way for performing LF deblurring as a single 2D blind deblurring (to estimate the common motion) and parallelizable *low-cost* 2D non-blind deblurring of multiple images. Our approach overcomes several major drawbacks of the state-of-the-art, such as heavy computational cost, ability to deblur *only* low-resolution LFs, and GPU-processing. Unlike the state-of-the-art, our model realistically captures refraction effects of lens, and works for wide-angle scenarios and irregular ego-motion as well. As LF cameras continue to evolve with higher resolutions, our divide and conquer strategy will be invaluable for full-resolution deblurring.

Acknowledgement: We thank the reviewers for their valuable comments. The first author thanks Sunil Kumar for his help in decoding Lytro files and running some comparison methods, and the Indian MHRD for student scholarship.

References

- [1] E. H. Adelson and J. Y. Wang. Single lens stereo with a plenoptic camera. *IEEE transactions on pattern analysis and machine intelligence*, 14(2):99–106, 1992. **1**
- [2] D. G. Dansereau, A. Eriksson, and J. Leitner. Richardson-lucy deblurring for moving light field cameras. *CVPR workshop on Light Fields for Computer Vision (CVPR:LF4CV)*, 2017. **2**
- [3] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1027–1034, 2013. **2**
- [4] D. G. Dansereau, G. Schuster, J. Ford, and G. Wetzstein. A wide-field-of-view monocentric light field camera. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. **1, 7**
- [5] A. Gupta, N. Joshi, L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *ECCV '10: Proceedings of the 10th European Conference on Computer Vision*, 2010. **1**
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. **4, 6**
- [7] M. R. Hatch. *Vibration simulation using MATLAB and ANSYS*. CRC Press, 2000. **7**
- [8] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 607–614. IEEE, 2010. **1, 2, 6**
- [9] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell. A category-level 3d object dataset: Putting the kinect to work. In *Consumer Depth Cameras for Computer Vision*, pages 141–165. Springer, 2013. **7**
- [10] M. Jin, P. Chandramouli, and P. Favaro. Bilayer blind deconvolution with the light field camera. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 10–18, 2015. **1, 2, 4**
- [11] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. 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. **1, 7**
- [12] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 233–240. IEEE, 2011. **1, 7, 8**
- [13] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang. A comparative study for single image blind deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1701–1709. IEEE, 2016. **7**
- [14] Z. Li, Z. Xu, R. Ramamoorthi, and M. Chandraker. Robust energy minimization for brdf-invariant shape from light fields. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. **1**
- [15] M. M. Mohan, A. Rajagopalan, and G. Seetharaman. Going unconstrained with rolling shutter deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4010–4018, 2017. **1, 6**
- [16] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a handheld plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005. **1, 2, 4, 6, 7**
- [17] J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1628–1636. IEEE, 2016. **1, 3, 4, 6, 7, 8**
- [18] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on image processing (TIP)*, 15(2):430–444, 2006. **7**
- [19] H. R. Sheikh, A. C. Bovik, and G. 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. **7**
- [20] P. P. Srinivasan, R. Ng, and R. Ramamoorthi. Light field blind motion deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. **1, 2, 3, 4, 7, 8**
- [21] S. Su and W. Heidrich. Rolling shutter motion deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1529–1537. IEEE, 2015. **1**
- [22] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 673–680, 2013. **1, 7**
- [23] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *International journal of computer vision (IJCV)*, 98(2):168–186, 2012. **1, 3, 6**
- [24] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 2017. **1**
- [25] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu. Snapshot hyperspectral light field imaging. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. **1**
- [26] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1107–1114. IEEE, 2013. **1, 3, 4, 6**