Going Unconstrained with Rolling Shutter Deblurring

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

Gunasekaran Seetharaman
U.S. Naval Research Laboratory
guna.seetharaman@nrl.navy.mil

Abstract

Most present-day imaging devices are equipped with CMOS sensors. Motion blur is a common artifact in hand-held cameras. Because CMOS sensors mostly employ a rolling shutter (RS), the motion deblurring problem takes on a new dimension. Although few works have recently addressed this problem, they suffer from many constraints including heavy computational cost, need for precise sensor information, and inability to deal with wide-angle systems (which most cell-phone and drone cameras are) and irregular camera trajectory. In this work, we propose a model for RS blind motion deblurring that mitigates these issues significantly. Comprehensive comparisons with state-of-the-art methods reveal that our approach not only exhibits significant computational gains and unconstrained functionality but also leads to improved deblurring performance.

1. Introduction

CMOS is winning the camera sensor battle as it offers advantages in terms of extended battery life, lower cost and higher frame rate, as compared to the conventional CCD sensor [14]. Nevertheless, the annoying effect of motion blur that affects CCD cameras prevails in common CMOS rolling shutter (RS) cameras too (Fig. 1), except that it manifests in a different form [24].

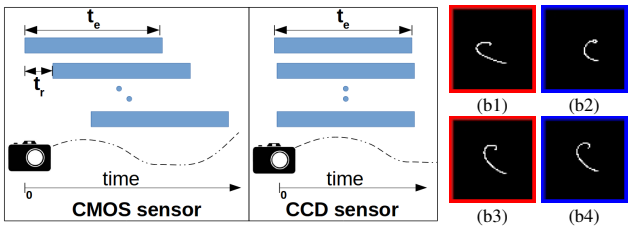
The problem of blind motion deblurring (BMD) – *i.e.*, recovery of both the clean image and underlying camera motion from a single motion blurred image – is an extensively studied topic for CCD cameras. Gupta *et al*. [4] proposed a 3D approximation for general 6D camera pose-space by considering only inplane translations and rotations, while Whyte *et al*. [27] considered only 3D rotations. Kohler *et al*. [12] showed that both these 3D models [27, 4] are good approximations to general pose-space. To reduce the ill-posedness of BMD, a recent trend is to introduce novel priors. Some representative works in this direction include natural image priors such as L_0 sparsity [28], total variation [18], and dark channel prior [16]; and ego-motion priors that include Tikhonov regularization [7], and sparsity

Input patch [24] {122 s, 44 s} Ours {**22** s, **1** s}

Figure 1. Rolling shutter motion deblurring: Comparison with state-of-the-art method [24]. Also given is the time taken for each ego-motion update and latent image update, respectively.

[27, 4] and continuity [4] in pose-space. Another research direction in BMD is towards reducing computational complexity. Cho and Lee [1] address this by utilizing the FFT for space invariant blur. Hirsch *et al*. [7, 8, 5] extend it to the space-variant case by approximating motion blur as space invariant over small image-patches, and show competitive quality with significant speed up.

However, the aforementioned deblurring methods proposed for CCD cameras are not applicable to CMOS-RS [24] since the RS motion blur formation is *strikingly different* as illustrated in Fig. 2(a). CCD camera uses a global shutter (GS), whereas CMOS cameras predominantly come with an electronic RS. In contrast to GS in which all sensor elements integrate light over the same time window (or experience the same camera motion), each sensor row in RS integrates over different time window, and thus a single camera motion does *not* exist for the entire image. To the best of our knowledge, *only* three works specifically address motion deblurring in RS cameras – [26] for depth camera videos, [10] for hardware assisted deblurring, and the BMD method of [24]. Tourani *et al*. [26] use feature matches between *depth maps* to timestamp parametric ego-motion. However, they require *multiple* RGBD images as input. Also, blurred RGB images, unlike depth maps, lack sufficient feature matches for reliable ego-motion estimation, which limits their functionality [26]. Hu *et al*. [10] use smartphone inertial sensors for timestamping and is thus a non-blind approach. Furthermore, it is device-specific and the cumulative errors from noisy inertial sensors and calibration govern deblurring performance [10].



(a) (b)

Figure 2. (a) Working principle of CMOS-RS and CCD sensors. (b) Effect of inplane rotation for a wide-angle system: (b1-b2) Two blur kernels (or PSFs) with inplane rotation and (b3-b4) corresponding PSFs *without* it. Note the variation in shape of the PSFs.

The current state-of-the-art RS-BMD [24] eliminates device-specific constrains of [10, 26], and estimates timestamped ego-motion solely from image intensities. However, the method is limited to parametric ego-motion derived specifically for hand-held blur. This renders it difficult to handle blur due to moving/vibrating platforms, such as in robotics, drones, street view cars *etc*. Also, to provide a good initialization, [24] *discards* inplane rotations and this precludes it from dealing with wide-angle systems [24]. The reason is illustrated in Fig. 2(b) with two different PSFs generated by a real ego-motion (using [12]) in a wide-angle system with and *without* the inplane rotation, which clearly reveals the inefficacy of their approximation. Wide-angle systems provide a larger field-of-view as compared to narrow-angle lenses, an important setting in most DSLR cameras, mobile phones and drones (supporting data is provided in the supplemental material). Another significant limitation of current RS deblurring methods [24, 10, 26] is their huge computational load. Moreover, methods [24, 10] require as input *precise* sensor timings t_r and t_e during image capture in order to fragment the estimated ego-motion corresponding to each image-row (see Fig. 2(a)). Other RS related works include RS super-resolution [19], RS image registration [20], RS structure from motion [11], *etc*.

In this paper, we propose an RS-BMD method that not only delivers excellent deblurring quality but is also computationally very efficient (see Fig. 1). It works by leveraging a generative model for RS motion blur (different from the one commonly employed), and a prior to disambiguate multiple solutions during inversion. Deblurring with our scheme not only relaxes the constraints associated with current methods, but also leads to an efficient optimization framework. Our main contributions are summarized below.

- Our method overcomes some of the major drawbacks of the state-of-the-art method [24], including inability to handle full 3D rotations (or wide-angle systems) and irregular ego-motion, and the need for sensor data.
- We extend the computationally efficient filter flow (EFF) framework that is commonly employed in CCD deblurring [8, 7] to RS deblurring. Relative to [24], we achieve a speed-up by a factor of *at least* eight.
- Ours produces state-of-the-art RS deblurring results for narrow- as well as wide-angle systems and under arbitrary ego-motion, all of these *sans* sensor timings.

2. RS Motion Blur Model

In this section, we discuss the generative model for RS motion blur. As mentioned earlier, the entire image in CCD (or GS) cameras experiences the *same* ego-motion. Thus the motion blurred image $\mathbf{B} \in \mathbb{R}^{M \times N}$ is generated by integrating the images seen by the camera along its trajectory during the exposure duration $[0, t_e]$ [24]. It is given by

$$\mathbf{B} = \frac{1}{t_e} \int_0^{t_e} \mathbf{L}^{\mathbf{p}(t)} dt, \tag{1}$$

where $\mathbf{p}(t_0)$ represents the general 6D camera pose at time instant t_0, $\mathbf{L}^{\mathbf{p}(t_0)}$ is the latent image $\mathbf{L}$ transformed according to the pose $\mathbf{p}(t_0)$, and t_e is the shutter speed.

In contrast, each RS sensor-row experiences *different* ego-motion due to staggered exposure windows (Fig. 2(a)). Unlike CCD, we cannot associate a global warp for the entire latent image $\mathbf{L}$, but need to consider each row separately. Image row $\mathbf{B}_i$ (subscript i indicates ith row) of an RS blurred image $\mathbf{B} = [{\mathbf{B}_1}^T \, {\mathbf{B}_2}^T \cdots \, {\mathbf{B}_M}^T]^T$ is given by

$$\mathbf{B}_i = \frac{1}{t_e} \int_{(i-1)\cdot t_r}^{(i-1)\cdot t_r + t_e} \mathbf{L}_i^{\mathbf{p}(t)} dt \quad : i \in \{1, 2, \cdots, M\}, \tag{2}$$

where $\mathbf{L}_i^{\mathbf{p}(t_0)}$ is the ith row of the transformed image $\mathbf{L}^{\mathbf{p}(t_0)}$, t_e is the shutter speed or row-exposure time in CMOS sensors, and t_r is the inter-row delay. All the current RS deblurring methods use a discretized form of Eq. (2) as the forward model, and we refer to this as temporal model.

An equivalent representation of Eq. (2) can be obtained by a weighted integration of the transformed image-rows over camera poses, where the weight corresponding to a transformed image-row with a specific pose determines the fraction of the row-exposure time (t_e) that the camera stayed in the particular pose. This is given by

$$\mathbf{B}_i = \int_{\mathbf{P}} w_i'(\mathbf{p}) \cdot \mathbf{L}_i^{\mathbf{p}} d\mathbf{p} \quad : i \in \{1, 2, \cdots, M\}, \tag{3}$$

where $\mathbf{P}$ is the pose-space and $w_i'(\mathbf{p_0})$ is the weight corresponding to the transformed row $\mathbf{L}_i^{\mathbf{Po}}$. Unlike existing RS deblurring works, we employ the second model and discretize the pose-space in Eq. (3). We consider the discretization step-size to be such that there is less than one pixel displacement between adjacent poses. Thus Eq. (3) reduces to

$$\mathbf{B}_i = \sum_{\mathbf{p} \in \mathbb{P}} w_i(\mathbf{p}) \cdot \mathbf{L}_i^{\mathbf{p}} \quad : i \in \{1, 2, \cdots, M\}, \tag{4}$$

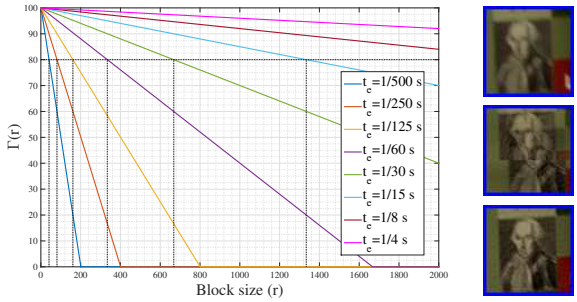


Figure 3. Left – Percentage pose-overlap Γ over block-size r for standard CMOS-RS shutter speed (t_e) & an inter-row delay (t_r) of $1/100$ ms, along with optimal block-size. Right – Top shows a blurred patch from Fig. 7(a)(top), and middle and bottom patches show the deblurred results *without* and with our RS prior.

where $\mathbb{P}$ is the discretized pose-space $\mathbf{P}$, and the discrete weight $w_i(\mathbf{p_0})$ is the summation of all the continuous-weights $w_i'(\mathbf{p})$ for all $\mathbf{p}$ that lie in the half step-size neighbourhood of pose $\mathbf{p_0}$. We identify the weights $w_i(\mathbf{p})$ as the motion density function (MDF), as in [4]. We further modify Eq. (4) based on an important observation derived from typical CMOS sensor settings.

Observation: *In RS motion blurred images, there exists an $r_b : 1 << r_b \leq M$, such that any block of contiguous rows with size less than or equal to r_b will have substantial camera-pose overlap.*

In RS sensors, the fraction of camera-pose overlap in r contiguous rows is equal to the fraction of the time shared among those rows. Thus, from Fig. 2(a), the percentage camera-pose overlap Γ in a block of r rows is obtained as

$$\Gamma(r) = \max\left(\frac{t_e - (r-1)\cdot t_r}{t_e}, 0\right)\cdot 100. \qquad (5)$$

In Fig. 3(left), we plot $\Gamma(r)$ for varying t_e and a fixed t_r of $1/100$ ms – a typical CMOS sensor has standardized t_e as $\{1/1000\,\text{s}, 1/500\,\text{s}, \cdots 1\,\text{s}\}$, and t_r in the range $1/200$ ms to $1/25$ ms [3]. It is evident from the figure that such a block-wise segregation is possible for these standard settings with camera-pose overlap of almost 80%. We do note that for faster shutter speed (*e.g.*, $t_e < 1/250$ s) r_b can be close to one; but for that setting motion blur will be negligible.

Based on this observation, we approximate each non-intersecting block of r_b rows that have substantial camera-pose overlap to be governed by an individual MDF. We will show that this approximation is reasonable for RS motion blurred images. Thus our forward RS motion blur model is given by

$$\mathbf{B_i} = \sum_{\mathbf{p}\in\mathbb{P}} w_{\mathbf{i}}(\mathbf{p})\cdot \mathbf{L_i^p} \quad : \mathbf{i} \in \{1, 2, \cdots, n_b\}, \qquad (6)$$

where $n_b = M/r_b$ is the total number of blocks, the blurred image $\mathbf{B}$ has structure $[\mathbf{B_1}^T, \mathbf{B_2}^T, \cdots \mathbf{B_{n_b}}^T]$ with $\mathbf{B_i}$ as the $\mathbf{i}$th block (bold subscript represents block), $w_{\mathbf{i}}(\mathbf{p})$ is the approximated MDF of $\mathbf{i}$th block, and $\mathbf{L_i^{Po}}$ is the $\mathbf{i}$th block of the transformed image $\mathbf{L}$ with pose $\mathbf{p_0}$. (Note that for $r_b = M$, Eq. (6) reduces to CCD motion blur model [16, 28, 27, 7, 4]). We identify Eq. (6) as our pose-space model. This is unlike the temporal model of [24] which constrains the motion model to be parametric. Therefore, our model can accommodate different kinds of motion trajectories including camera shake and vibrations.

3. RS Deblurring

We formulate a maximum a posteriori (MAP) framework for estimation of both the latent image and the block-MDFs. In this section, we introduce a new prior for MDF to enable RS motion deblurring.

A direct MAP framework for unknown $\theta = \{\mathbf{L}, \mathbf{w_i} : 1 \leq \mathbf{i} \leq n_b\}$ is given as

$$\hat{\theta} = \min_{\theta} \sum_{\mathbf{i}=1}^{n_b} \|(\mathbf{B_i} - \sum_{\mathbf{p}\in\mathbb{P}} w_{\mathbf{i}}(\mathbf{p})\cdot \mathbf{L_i^p}\|_2 + \lambda_1\|\nabla\mathbf{L}\|_1 + \lambda_2 \sum_{\mathbf{i}=1}^{n_b} \|\mathbf{w_i}\|_1, \qquad (7)$$

where $\mathbf{w_i}$ is the vector containing weights $w_{\mathbf{i}}(\mathbf{p})$ for poses $\mathbf{p} \in \mathbb{P}$ and $\nabla\mathbf{L}$ is the gradient of $\mathbf{L}$. We assume that the optimal block-size r_b is known (we relax this subsequently in section 4.4). The first term in the objective is data fidelity that enforces our forward blur model of Eq. (6). To reduce ill-posedness, we too enforce a sparsity prior on the image-gradient following [27, 4]. We also impose a sparsity prior on the MDF weights since a camera can transit over only few poses in $\mathbb{P}$ during exposure. In the literature on CCD deblurring (*i.e.*, $n_b = 1$) it is well-known that the objective in Eq. (7) is biconvex, *i.e.*, it is individually convex with respect to the latent image and MDF, but non-convex overall; and convergence to a local minima is ensured with alternative minimization of MDF and latent image [27, 4, 1]. However, RS sensors introduce a different challenge if Eq. (7) is directly considered (Fig. 3(right), middle patch).

Claim 1: *For an RS blurred image, there exist multiple solutions for the latent image-MDF pair in each individual image-block. They satisfy the forward model in Eq. (6) and are consistent with the image and MDF prior in Eq. (7).*

Before giving a formal proof, we attempt to provide some intuition. Considering *only* inplane rotations, Fig. 4 illustrates a multiple-solution case where one latent image-block (of solution-pair $\{\mathbf{L_i'}, \mathbf{w_i'}\}$) is rotated anti-clockwise and the second image-block (of solution-pair $\{\mathbf{L_i''}, \mathbf{w_i''}\}$) is rotated clockwise, but both result in the *same* input image-block $\mathbf{B_i}$. This can also be visualized as a natural escalation of the notion of shift-ambiguity in patch-wise PSF estimates [17] all the way to block-wise MDFs.

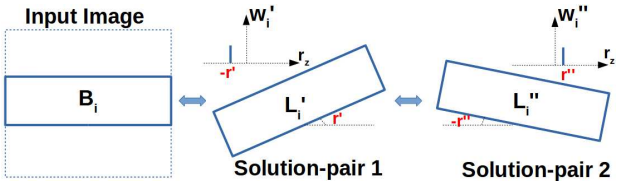


Figure 4. Illustration of block-wise latent image-MDF pair ambiguity for a single inplane rotation (only 1D pose-space). Both solution-pairs 1 and 2 result in the same blurred block $\mathbf{B_i}$.

Proof: Let $\mathbf{B_i}$ be an RS blurred block formed by latent image $\mathbf{L}$ and MDF $\mathbf{w_i}$ through Eq. (6). We form a second RS blurred block $\mathbf{B'_i}$ by considering a *nonzero* pose $\mathbf{p_0} \in \mathbb{P}$ as

$$\mathbf{B'_i} = \sum_{\mathbf{p'} \in \mathbb{P}} w_{\mathbf{i}}(\mathbf{p_0} + \mathbf{p'}) \cdot \mathbf{L_i^{p_0+p'}}, \tag{8}$$

where $\mathbf{L_i^{p_0+p'}}$ is the ith block of the transformed version of $\mathbf{L^{p_0}}$ with pose $\mathbf{p'}$, and $w_{\mathbf{i}}(\mathbf{p_0} + \mathbf{p'})$ is obtained by shifting $w_i(\mathbf{p'})$ with a negative offset of $\mathbf{p_0}$. Even though the latent image-MDF pairs for $\mathbf{B_i}$ and $\mathbf{B'_i}$ are different, *i.e.*, $\{\mathbf{L}, w_{\mathbf{i}}(\mathbf{p})\}$ in Eq. (6), and $\{\mathbf{L^{p_0}}, w_{\mathbf{i}}(\mathbf{p_0} + \mathbf{p})\}$ in Eq. (8), we shall prove that both $\mathbf{B_i}$ and $\mathbf{B'_i}$ are equal. Construct a set $\mathbb{S}_{\mathbb{B}i}$ with elements as *all* individual additive components of Eq. (6) that add up to get $\mathbf{B_i}$. Similarly, form set $\mathbb{S}'_{\mathbb{B}i}$ with *all* additive components of Eq. (8). Any element in $\mathbb{S}_{\mathbb{B}i}$ is represented as a singleton $\{w_{\mathbf{i}}(\mathbf{p}) \cdot \mathbf{L_i^p}\}$ with $\mathbf{p} \in \mathbb{P}$. The same element is present in $\mathbb{S}'_{\mathbb{B}i}$, *i.e.*, at $\mathbf{p'} = -\mathbf{p_0} + \mathbf{p}$ in Eq. (8), which implies $\mathbb{S}_{\mathbb{B}i} \subseteq \mathbb{S}'_{\mathbb{B}i}$. Similarly by considering $\mathbf{p} = \mathbf{p_0} + \mathbf{p'}$ in Eq. (6), it follows that $\mathbb{S}'_{\mathbb{B}i} \subseteq \mathbb{S}_{\mathbb{B}i}$. Since $\mathbb{S}_{\mathbb{B}i} \subseteq \mathbb{S}'_{\mathbb{B}i}$ and $\mathbb{S}'_{\mathbb{B}i} \subseteq \mathbb{S}_{\mathbb{B}i}$, both the sets are equal, and so are $\mathbf{B_i}$ and $\mathbf{B'_i}$. Also, as the latent images $\mathbf{L}$ and $\mathbf{L^{p_0}}$ are related by a global warp, the sparsity in gradient domain (*i.e.*, image prior in Eq. (7)) is valid for both. Since both the MDFs have equal weight distribution, the sparsity in weights (*i.e.*, MDF prior) is also identical for both. Hence proved. ■

If we consider Eq. (7) *alone* for RS deblurring, this ambiguity can cause the latent image portion of *individual* block to transform *independently* (see Fig. 4). This can result in an erroneous estimate of the deblurred image, where the latent image portions corresponding to different blurry blocks are incoherently combined (Fig. 3(right)). To address this issue, we introduce an additional prior on the MDFs as

$$\text{prior}(\mathbf{w}) = \sum_{\mathbf{i}=1}^{n_b} \sum_{\mathbf{j}>\mathbf{i}}^{n_b} \|\Gamma\left(r_b(\mathbf{j}-\mathbf{i}+1)\right) \cdot (\mathbf{w_i} - \mathbf{w_j})\|_2^2, \tag{9}$$

where $\Gamma\left(r_b \cdot (\mathbf{j}-\mathbf{i}+1)\right)$ is the percentage overlap (Eq. (5)) of all groups of blocks between (and including) the $\mathbf{i}$th and $\mathbf{j}$th block, and $\mathbf{w}$ is a vector obtained by stacking all the unknown MDFs $\{\mathbf{w_i} : 1 \leq \mathbf{i} \leq n_b\}$. This prior restricts drifting of MDFs between neighbouring blocks (*i.e.*, high cost), but allows MDFs to change between distant blocks (*i.e.*, low cost). It also serves to impart an additional dependency between block MDFs which Eq. (7) does not possess. This helps to reduce the ill-posedness of ego-motion estimation.

Claim 2: *The prior in Eq. (9) is a convex function in* $\mathbf{w}$*, and can be represented as a norm of matrix vector multiplication,* i.e.*, as* $\|\mathbf{Gw}\|_2^2$*, with sparse* $\mathbf{G}$. (Proof is given in the supplementary material.) Thus inclusion of the prior does *not* alter the biconvexity of Eq. (7) (which is necessary for convergence), and paves the way for efficient implementation (as we shall see in section 4.2). We identify Eq. (9) as our proposed RS prior.

4. Model and Optimization

State-of-the-art CCD-BMD methods [16, 28, 27, 1] work by alternative minimization (AM) of MDF and latent image over a number of iterations in a scale-space manner, (*i.e.*, AM proceeds from coarse to fine image-scale in order to handle large blurs). As we shall see shortly, this requires generation of blur numerous times. Efficiency of the blurring process is a major factor that governs computational efficiency of a method. In this section, we first discuss how our pose-space model allows for an efficient process for RS blurring (analogous to CCD-EFF [7, 8]). Next, we elaborate on our AM framework, and eventually relax the assumption of the need for sensor information.

4.1. Efficient Filter Flow for RS blur

Following [8], we approximate motion blur in individual *small* image patches as space invariant convolution with different blur kernels. We represent this as

$$\mathbf{B} = \sum_{k=1}^{R} \mathbf{C_k^\dagger} \cdot \left\{ a^{(k,b(k))} * (\mathbf{C_k} \cdot \mathbf{L}) \right\}, \tag{10}$$

where R is the total number of overlapping patches in latent image $\mathbf{L}$, $b(k)$ is a function which gives the index of the block to which the major portion of the k^{th} patch belongs (*i.e.*, $b(k) \in \{\mathbf{1}, \mathbf{2}, \cdots \mathbf{n_b}\}$), $\mathbf{C_k} \cdot \mathbf{L}$ is a linear operation which extracts the kth patch from $\mathbf{L}$, and $\mathbf{C_k^\dagger}$ inserts the patch back to its original position with a windowing operation. $a^{(k,b(k))}$ represents the blur kernel which when convolved with the kth latent image-patch creates blurred patch. Considering $b(k)$ as $\mathbf{j}$, we can write

$$a^{(k,b(k))} = \sum_{\mathbf{p} \in \mathbb{P}} w_{\mathbf{j}}(\mathbf{p}) \cdot \delta_k(\mathbf{p}), \tag{11}$$

where $\delta_k(\mathbf{p})$ is a shifted impulse obtained by transforming with pose $\mathbf{p}$ an impulse centered at the kth patch-center. Intuitively, the blur kernel at patch k due to an arbitrary MDF is the superposition of the $\delta_k(\mathbf{p})$s generated by it. Since $\delta_k(\mathbf{p})$ is independent of the latent image and the MDF, it needs to be computed *only once*, and can be subsequently used to create the blur kernel in patch k for any

image. Thus, given a latent image $\mathbf{L}$ and MDF of each block, our blurring process first computes kernels in R patch-centres using the precomputed $\delta_k(\mathbf{p})$ (Eq. 11), convolves them with their corresponding latent-image patches and combines them to form the RS blurred image (Eq. (10)). We carry out convolution using the *efficient* FFT. Note that the CCD-EFF is a special case of Eq. (10) under identical MDFs ($\mathbf{w_i} = \mathbf{w}\ \forall \mathbf{i}$) or the single block case ($n_b = 1$).

We next discuss our AM at the finest level. The same procedure is followed at coarser levels and across iterations.

4.2. Ego-Motion Estimation

The objective of this step is to estimate the ego-motion at iteration $d+1$ (*i.e.*, $\mathbf{w}^{d+1}$) given the latent image estimate at iteration d (*i.e.*, $\mathbf{L}(d)$). We frame our MDF objective function in the gradient domain for faster convergence and to reduce ill-conditionness [28, 27, 1, 7]. We give it as

$$\mathbf{w}^{d+1} = \arg\min_{\mathbf{w}} \|\mathbf{F}\mathbf{w} - \nabla\mathbf{B}\|_2^2 + \alpha\|\mathbf{G}\mathbf{w}\|_2^2 + \beta\|\mathbf{w}\|_1, \quad (12)$$

where the information of the gradient of $\mathbf{L}(d)$ is embedded in blur matrix $\mathbf{F}$, $\nabla\mathbf{B}$ is the gradient of $\mathbf{B}$, and $\|\mathbf{G}\mathbf{w}\|_2$ is the prior we introduce for RS blur (Eq. 9). We further simplify the objective in Eq. (12) by separating out the sparsity prior as a constraint and taking the derivative. This yields

$$\mathbf{w}^{d+1} = \arg\min_{\mathbf{w}} \|(\mathbf{F}^T\mathbf{F} + \alpha\mathbf{G}^T\mathbf{G})\mathbf{w} - \mathbf{F}^T\nabla\mathbf{B}\|_2^2 \quad \text{such that } \|\mathbf{w}\|_1 \leq \beta'. \quad (13)$$

A major advantage of our pose-space model over the temporal model of Eq. (2) is that we can formulate ego-motion estimation as in Eq. (13). This is LASSO [25] of the form $\arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 : \|\mathbf{x}\|_1 \leq \gamma$, and has efficient solvers (LARS [2]). Since LASSO needs no initialization, we can account for inplane rotations too. This allows us to handle even wide-angle systems unlike [24], which discards it to eliminate the back-projection ambiguity of blur kernels in its initialization step.

Suppose a blurred image of size $M \times N$ and n_b number of MDFs of length l (*i.e.*, block-size $r_b = M/n_b$). Then the *dense* matrix $\mathbf{F}$ in Eq. (13) is n_b times larger compared to the CCD case. This escalates the memory requirement and computational cost for RS deblurring; *i.e.*, a naive approach to create $\mathbf{F}^T\mathbf{F}$ (with size $n_b \cdot l \times n_b \cdot l$) is to form a *large* matrix $\mathbf{F}$ of size $MN \times n_b \cdot l$ (where $MN >> n_b l$), and perform large-matrix multiplication. We avoid this problem by leveraging the block-diagonal structure of $\mathbf{F}$, and thus for $\mathbf{F}^T\mathbf{F}$, that is *specific* for RS blur. The jth column of the $\mathbf{i}$th block-matrix $\mathbf{F_i}$ of $\mathbf{F}$ (of size $r_b N \times l$) is formed by transforming $\nabla\mathbf{L}(d)$ with the pose of $w_\mathbf{i}(j)$, and vectorizing its $\mathbf{i}$th block. For this, we employ the RS-EFF. Since each $\mathbf{F_i}$ can be generated independently, we bypass creating $\mathbf{F}$, and instead directly arrive at the block-diagonal matrix $\mathbf{F}^T\mathbf{F}$, one diagonal-block at a time, with the jth block as $\mathbf{F_j}^T\mathbf{F_j}$. A similar operation is also done for $\mathbf{F}^T\nabla\mathbf{B}$. Since $\mathbf{G}$ is sparse, $\mathbf{G}^T\mathbf{G}$ in Eq. (13) can be computed efficiently [29].

4.3. Latent Image Estimation

Given the ego-motion at iteration $d+1$ (*i.e.*, $\mathbf{w}^{d+1}$), this step estimates the latent image $\mathbf{L}(d+1)$. Since ego-motion estimation is based on image gradients (Eq. (13)), only the latent-image gradient information needs to be correctly estimated, as pointed out in [1, 7, 27]. This eliminates the use of computationally expensive image priors in the alternative minimization step. We obtain the latent image by inverting the forward blurring process in Eq. (10), *i.e.*,

$$\mathbf{L}(d+1) = \sum_{k=1}^{R} \mathbf{C_k}^{\dagger} \cdot \mathbb{F}^{-1}\left(\frac{1}{\mathbb{F}(a^{(k,b(k))})} \odot \mathbb{F}(\mathbf{C_k} \cdot \mathbf{B})\right), \quad (14)$$

where $a^{(k,b(k))}$ is generated using $\mathbf{w}^{d+1}$, $\mathbb{F}$ and $\mathbb{F}^{-1}$ are the forward and inverse DFT, respectively, and $\odot$ is a point-wise multiplication operator that also suppresses unbounded-values. We combine patches using Bartlett-Hann window that tapers to zero near the patch boundary. It has 70% overlap for patches that span adjacent blocks (to eliminate the effect of sudden changes in MDFs), and 50% for the rest. It is important to note that explicit block-wise segregation of blurred image is employed *only* for MDF estimation (to create $\mathbf{F}^T\mathbf{F}$ in Eq. (13)), and *not* for latent image estimation where the estimated MDFs are utilized *only* to project PSFs in overlapping patches, akin to CCD-BMD [27, 7, 8]. From a computational perspective, this is equivalent to extracting each patch of the blurred image, deconvolving it with the corresponding blur kernel (created using Eq. (11)) with FFT acceleration, and combining the deconvolved patches to form the updated latent image. For the final iteration (in the finest level), instead of FFT inversion (as in Eq. (14)), we adopt Richardson-Lucy deconvolution [15] which considers natural image-priors.

4.4. Selection of Block-Size

In section 3, we had assumed the availability of sensor timings t_r and t_e to optimally segregate image-blocks (using $\Gamma(r)$ in Eq. (5)) and to derive the RS prior (through Γ in Eq. (9)). In this section, we quantify camera-pose overlap and relax the need for sensor timings.

To analyse the effect of block-size r_b, we conducted an experiment using real camera trajectories from the dataset of [12] with CCD blur. Since all the rows would experience a common trajectory, *ideally* all block MDFs should match irrespective of the chosen number of blocks. We estimate MDFs *without* using the RS prior ($\alpha = 0$ in Eq. (13)), align their centroids, and use individual MDF to compute PSFs in all R patches. The PSFs are then correlated with the ground-truth PSFs using the kernel similarity metric in

[9] (centroid alignment of MDFs was done since correlation cannot handle arbitrary rotation between PSFs). We plot in Fig. 5(left) average kernel similarity and time taken for AM steps at the finest level for different block-sizes. A key observation is that as the block-size r_b falls below 134 (*i.e.*, $n_b \geq 6$ for an 800×800 image dimension), kernel similarity drops significantly. This ineffectiveness for smaller blocks is due to lack of sufficient image gradients *within* individual blocks for MDF estimation. Also, the computation time increases as the block-size is reduced. These factors translate to a pose-intersection of 80%, which gives a reliable block-size for typical CMOS settings (Fig. 3(left)).

Apart from sensor timings, camera ego-motion can also influence optimal block-size, *e.g.*, if the camera moves slowly during the exposure of few blocks, merging them as a single block can reduce the number of unknown MDFs. Hence, we allow for variable-sized blocks. We can segregate image-blocks *without* using sensor information as described next. Given a blurred image, we convert the image to a coarse level ($M_0 \times N_0$), and estimate MDFs *without* the RS prior assuming uniform block-size of r_0. For each MDF, we find kernel similarity with the neighbouring MDFs (as described earlier). Adjacent blocks with kernel similarity greater than 0.8 are combined until no such merging is possible. The resultant block-sizes at the coarse level multiplied by the upsampling factor to the finest image-level are considered for final segregation. For the RS prior, since the pose-overlap $\Gamma(r)$ in Eq. (5) is parametrised by a single unknown (which is t_r/t_e), we solve for t_r/t_e (*without* requiring it from RS sensor) assuming the number of rows (r) for the smallest segregated block as having 80% Γ.

4.5. Computational Aspects

First, our pose-space model (Eq. (6)) performs RS blurring efficiently as already discussed in section 4.1. In contrast, the blurring process adopted in current RS deblurring methods is relatively quite expensive. That is, given a latent image $\mathbf{L}$ and temporal ego-motion (with N_t temporal bins), an RS blurred image is created by N_t *individual* transformations of $\mathbf{L}$ – each using individual warping and bilinear interpolation *over all image-locations* – and the rows are combined using sensor timings (Eq. (2) and Fig. 2). Second, our pose-space model together with the RS prior is amenable to the very-efficient LARS framework (Eq. (13)). In contrast, because of the parametric model (polynomial in [24] and splines in [26]) for ego-motion in temporal domain, those methods need to employ non-linear optimization (Eq. (9) in [24] and Eq. (8) in [26]), which is much more expensive than LARS [2].

Third, our method employs patch-wise deblurring leveraging the very efficient FFT (rather than optimizing the full image with prior). In contrast, as stated under Eq. (15) of [24], for *every* iteration it must optimize the high-dimensional latent image with a costly prior as

$$\mathbf{L}(d+1) = \arg\min_{\mathbf{L}} \|\mathbf{XL} - \mathbf{B}\|_2^2 + \|\nabla \mathbf{L}\|_1, \qquad (15)$$

where $\mathbf{L}$ and $\mathbf{B}$ are vectorized latent and blurred images of size $MN \times 1$, respectively, and $\mathbf{X}$ is a sparse matrix of size $MN \times MN$. Generating $\mathbf{X}$ using the expensive forward model and optimization with a costly prior is a serious computational bottleneck for [24].

Image dimension	Ego-motion estimation time (s)		Latent image estimation time (s)	
ht. $\times$ wd.	[24]	Ours	[24]	Ours
800×800	216.01	**29.58**	258.65	**1.44**
450×800	122.28	**22.48**	44.23	**1.30**
400×400	73.26	**10.34**	23.82	**0.62**

Table 1. Time comparisons with state-of-the-art [24].

To determine the computational gains of our proposal, we conducted experiments on variable-sized images. The average time taken for each AM step at finest level using `MATLAB` is listed in Table 1 (using the code of [24] from the author's website). It is clearly evident that our method offers significant computational gains. Note the improvement of ego-motion estimation from 216 to 30 seconds, and latent image estimation from 258 to 2 seconds for an 800×800 image. Also, observe the steep rise in computational cost for [24] with image size, unlike ours.

5. Experimental Results

In this section, we demonstrate that our method can handle both wide- and narrow-angle systems, arbitrary ego-motion, and RS as well as GS blurs. We used default parameters for all the competing methods. Our supplementary provides our system specification and the parameters used.

Datasets used: For quantitative evaluation, we created RS motion blurred images using hand-held trajectories from the benchmark dataset in [12]. We used focal length 29 mm (for wide-angle) and 50 mm (for narrow-angle), and sensor timings of $t_r = 1/50$ ms and $t_e = 1/50$ s (as in [24]). Vibration motion was taken from [6]. For real experiments, we considered individually the cases of RS narrow-angle, RS wide-angle and CCD blurs. For the narrow-angle case, we used the dataset in [24]. Since RS wide-angle configuration has not been hitherto addressed, we created an RS wide-angle blur dataset which contains images captured with iPhone 5S (focal length 29 mm). We also considered drone images from the internet characterizing irregular ego-motion. For CCD blur, we used the dataset in [16].

RS deblurring comparisons: We considered mainly the current RS-BMD state-of-the-art method [24] for evaluation. Since [10] requires inertial sensor data, and [26] uses

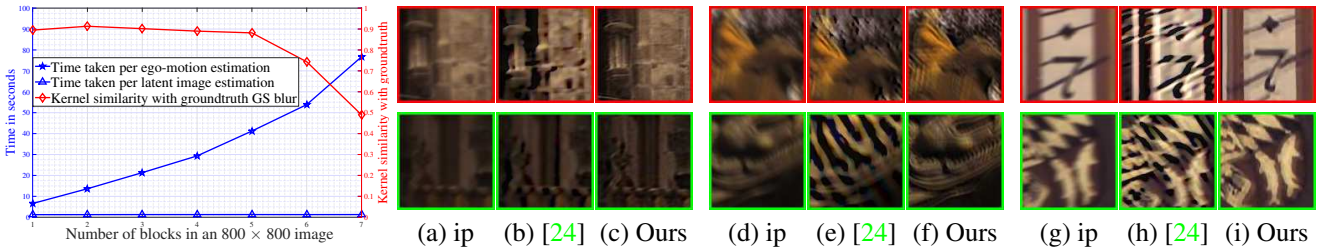


Figure 5. Left – Analysis on the effect of block-size. Right portion – Synthetic experiments with {IFC,VIF}: (a-c) For an RS wide-angle system ([24] - $\{1.31, 0.23\}$, Ours - $\{\mathbf{1.97}, \mathbf{0.36}\}$), (d-f) For vibratory motion in an RS system ([24] - $\{\ 0.49, 0.076\ \}$, Ours - $\{\mathbf{0.59}, \mathbf{0.086}\}$), and (g-i) For GS blur ([24] - $\{1.25, 0.19\}$, Ours - $\{\mathbf{2.11}, \mathbf{0.32}\}$). (Full images are provided in the supplementary.)

multiple RGBD images, these techniques do not address BMD and thus have been omitted for comparisons (also their codes are not available). To analyse the performance of CCD-BMD methods on CMOS data, we also tested with [28]. Since the space-variant (SV) code of [16] (the best CCD-BMD) is *not* available, we evaluated using [28] (the second best). For comparisons, we downloaded the codes from the websites of the authors of [24] and [28]. We provide several additional comparisons in the supplementary.

GS deblurring comparisons: Since the SV code of [16] is not available, we report results on the SV examples provided in [16]. We also give comparisons with other SV-BMD methods [28, 27, 4] in the supplementary.

Quantitative evaluation: As pointed out in a very recent comparative study of BMD [13], information fidelity criterion (IFC) [23] and visual information fidelity (VIF) [22] are important metrics for evaluating BMD methods (higher values are better). Thus we adopt the same. Also, we wish to highlight the observation in [21] that ringing artifacts in deblurring are mainly caused by ego-motion estimation error, which can be either due to inaccurate blur/ego-motion model or ineffectiveness of optimization. Hence, we also use ringing as an evaluation tool. For visual comparisons, we show patches from upper and lower image portions.

First, we consider a wide-angle system (29 mm and trajectory #39 [12]) in Figs. 5(a-c). As shown in Fig. 5(b), the result of [24] shows moderate ringing artifacts in the top patch (in the wall-linings and lantern), whereas residual blur exists in the lower patch (in the table structure). In contrast, our result (Fig. 5(c)) recovers fine details in both the patches and with *no* ringing artifacts. This reveals our method's ability to deal with wide-angle systems, unlike [24] which is designed for *only* narrow-angle systems.

Next we consider vibration ego-motion (Figs. 5(d-f)) that simulates a robotic system with feedback control [6]. Since [24] considers *only* narrow-angle systems, we also limit ourselves to narrow-angle setting (50 mm), for a fair comparison. We found that the polynomial model, as considered in [24], fails to approximate these vibrations (illustrations are given in the supplementary). It is evident from

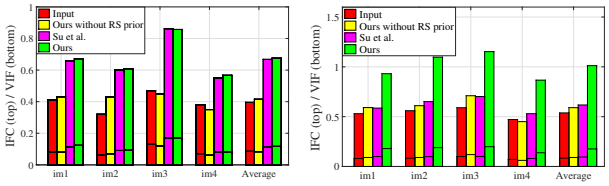


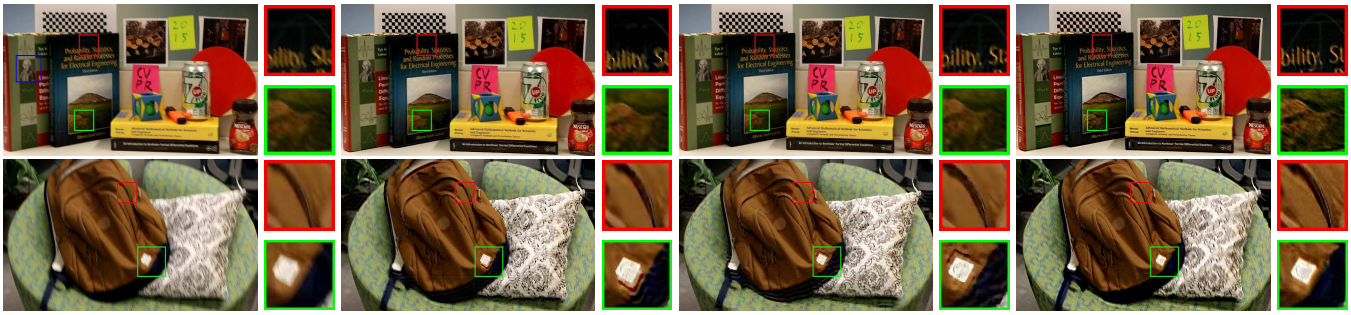
Figure 6. Quantitative evaluation on benchmark dataset [12] with RS settings. The performance of our method is comparable to that of [24] for narrow-angle systems but outperforms [24] for wide-angle systems; both *sans* RS timings t_r and t_e, unlike [24].

the results of Fig. 5(e) that the estimated polynomial model fits the initial portion of the trajectory well (top-patch is deblurred), but diverges for the later portion (heavy ringing in bottom patch). In contrast, our method gives good deblurred result uniformly (Fig. 5(f)), which underscores the importance of our non-parametric approach to ego-motion.

Finally, we evaluate the performance of RS BMD methods for GS deblurring in Figs. 5(g-i). Here also, we limit to narrow-angle systems (50 mm, trajectory #2 [12]). Either due to the ineffectiveness of polynomial approximation or initialization error, the result of [24] in Fig. 5(h) has moderate ringing artifacts with residual blurs. Our result in Fig. 5(i) reveals that our model generalizes to GS blur well, as compared to [24].

A detailed evaluation on dataset [12] for narrow- and wide-angle systems is given in Fig. 6. It clearly reveals that our method is either comparable to or better than [24] for narrow-angle systems. The performance of our method is strikingly superior for wide-angle systems. Note that all these are achieved *without* requiring t_r and t_e, unlike [24].

Real examples: In Fig. 7, we evaluate our method on real examples for RS narrow-angle systems using the dataset of [24]. The output of [24] (third column) contains residual blur and ringing artifacts compared to ours. Specifically, in the top-row, the characters in patch 1 and details in patch 2 are sharper in our output. In the second-row, the minute structures of the bag-zipper in patch 1 are restored well,



(a) Input (b) Xu *et al*. [28] (c) Su and Heidrich [24] (d) Ours

Figure 7. Comparisons for RS *narrow-angle* blur dataset of [24]. (a) Input, (b) CCD-BMD [28], (c) State-of-the-art RS-BMD [24], and (d) Ours. Our method provides negligible ringing artifacts and fine details, as compared to [24]. (Table 1(450×800 entry) gives the speed-up.)

(a) Input (b) Xu *et al*. [28] (c) Su and Heidrich [24] (d) Ours

Figure 8. Comparisons for RS *wide-angle* examples (first row - low-light case, second row - well-lit case, and third row - drone imaging). (a) Input, (b) CCD-BMD [28], (c) State-of-the-art RS-BMD [24], and (d) Ours. Ours consistently outperforms competing methods.

while ringing in patch 2 is negligible. In Fig. 8, we evaluate our algorithm for wide-angle systems (including irregular camera motion). The first row shows the case of low-light imaging, the second row is a well-lit scenario, and the third row is a drone image (irregular ego-motion). It is evident from the results of Fig. 8 that our method consistently delivers good performance over [24] in all the scenarios. The performance degradation of [24] in Fig. 8 may be attributed to its inability to handle wide-angle systems (first and second rows) and irregular ego-motion (third row). Figs. 7 and 8 (second column) also reveal the inability of CCD deblurring methods to handle blur in RS systems.

6. Concluding remarks

In this paper, we proposed a block-wise RS blur model for RS deblurring. We provided a detailed analysis of this model, and addressed invertibility issues using a computationally tractable convex prior. We also proposed an efficient filter flow framework that offers significant computational edge. Experiments reveal that our algorithm achieves state-of-the-art results in terms of deblurring quality as well as computational efficiency. Unlike existing RS deblurring methods, it can seamlessly accommodate wide- and narrow-angle systems, blur due to hand-held and irregular ego-motion, and GS as well as RS images; all *without* the need for sensor information. Our future work will consider incorporating effective priors such as those used in CCD cameras into the realm of RS deblurring.

Acknowledgement: A part of this work was supported by a grant from the Asian Office of Aerospace Research and Development, AOARD/AFOSR. The support is gratefully acknowledged. The results and interpretations presented in this paper are that of the authors, and do not necessarily reflect the views or priorities of the sponsor. The first author thanks Vijay Rengarajan for his help in improving the paper presentation, and the Indian MHRD for student scholarship.

References

[1] S. Cho and S. Lee. Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, page 145. ACM, 2009. 1, 3, 4, 5

[2] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004. 5, 6

[3] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar. Coded rolling shutter photography: Flexible space-time sampling. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2010. 3

[4] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *European Conference on Computer Vision (ECCV)*, pages 171–184. Springer, 2010. 1, 3, 7

[5] S. Harmeling, H. Michael, and B. Schölkopf. Space-variant single-image blind deconvolution for removing camera shake. In *Advances in Neural Information Processing Systems (NIPS)*, pages 829–837, 2010. 1

[6] M. R. Hatch. *Vibration simulation using MATLAB and ANSYS*. CRC Press, 2000. 6, 7

[7] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *IEEE International Conference on Computer Vision (ICCV)*, pages 463–470. IEEE, 2011. 1, 2, 3, 4, 5

[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, 4, 5

[9] Z. Hu and M.-H. Yang. Good regions to deblur. In *European Conference on Computer Vision (ECCV)*, pages 59–72. Springer, 2012. 6

[10] Z. Hu, L. Yuan, S. Lin, and M.-H. Yang. Image deblurring using smartphone inertial sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1855–1864. IEEE, 2016. 1, 2, 6

[11] E. Ito and T. Okatani. Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017. 2

[12] 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, 2, 5, 6, 7

[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] D. Litwiller. Ccd vs. cmos. *Photonics Spectra*, 35(1):154–158, 2001. 1

[15] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79:745, 1974. 5

[16] 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

[17] C. Paramanand and A. N. Rajagopalan. Non-uniform motion deblurring for bilayer scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1115–1122. IEEE, 2013. 3

[18] D. Perrone and P. Favaro. Total variation blind deconvolution: The devil is in the details. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2909–2916. IEEE, 2014. 1

[19] A. Punnappurath, V. Rengarajan, and A. Rajagopalan. Rolling shutter super-resolution. In *IEEE International Conference on Computer Vision (ICCV)*, pages 558–566. IEEE, 2015. 2

[20] V. Rengarajan, A. N. Rajagopalan, R. Aravind, and G. Seetharaman. Image registration and change detection under rolling shutter motion blur. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 2016. 2

[21] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *ACM transactions on graphics (TOG)*, volume 27, page 73. ACM, 2008. 7

[22] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on image processing (TIP)*, 15(2):430–444, 2006. 7

[23] 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

[24] 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, 2, 3, 5, 6, 7, 8

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

[26] S. Tourani, S. Mittal, A. Nagariya, V. Chari, and M. Krishna. Rolling shutter and motion blur removal for depth cameras. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5098–5105. IEEE, 2016. 1, 2, 6

[27] 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, 4, 5, 7

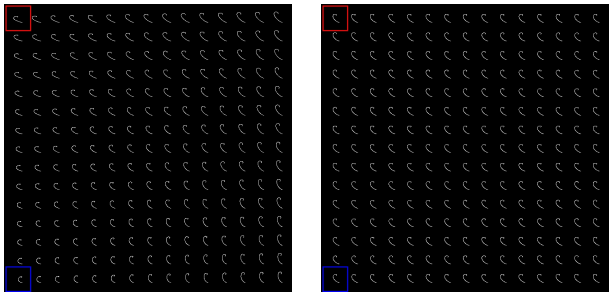
[28] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1107–1114. IEEE, 2013. 1, 3, 4, 5, 7, 8

[29] R. Yuster and U. Zwick. Fast sparse matrix multiplication. *ACM Transactions on Algorithms (TALG)*, 1(1):2–13, 2005. 5

Going Unconstrained with Rolling Shutter Deblurring (Supplementary Material)

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

Gunasekaran Seetharaman
U.S. Naval Research Laboratory
guna.seetharaman@nrl.navy.mil



(a) With inplane rotations (b) *Without* inplane rotations

Figure S1. Full set of PSFs of Fig. 2(b) illustrating the effect of inplane rotations for wide-angle setting (29 mm) using [6].

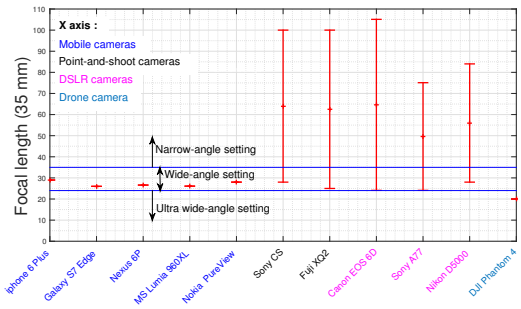


Figure S2. Focal lengths of some popular CMOS devices. Note the wide-angle setting predominant in cell-phone and drone cameras.

We begin by revisiting the problem motivation. This is followed by section S2 which contains our proof for claim 2 (section 3). Section S3 gives implementation details, and section S4 is devoted to additional evaluations.

S1. Problem Motivation (illustrative)

As mentioned in section 1 in the main paper, our method advances the state-of-the-art in RS deblurring, as it can deal with wide-angle configuration, unconstrained ego-motion and unconstrained shutter, *without* the need for timing information. Here, we further elaborate the significance of these problems that we have addressed in our work.

Unconstrained Focal-length: The PSFs provided in Fig. 2(b), which illustrates the importance of inplane rotations for wide-angle systems, is created using a focal length of 29 mm and real hand-held trajectory #39 in [6]. The full set of PSFs is provided in Fig. S1. We give in Fig. S2 focal-length settings of some popular CMOS imaging devices. It is clearly evident from the figure that wide-angle configurations are indeed important in photography (and predominant in cell-phones and drone cameras). However, the state-of-the-art RS-BMD [11] works only for narrow-angle settings. Hence, it is important to accommodate wide-angle settings.

Unconstrained Ego-motion: Even though a polynomial function can reasonably model human camera shake, RS blur also exists in images captured by drones, street view cars, *etc*., wherein the ego-motion is seldom regular [11]. Fig. S3 illustrates this fact with an under-damped response of a robotic system (which we employed in Figs. 5(d-f) using [4]). Also given is the approximation using a fourth order polynomial (as used in state-of-the-art RS-BMD [11]). From the plot it is clear that the polynomial model is *unable* to adequately capture the motion, thus underscoring the need for handling unconstrained ego-motion.

RS timing information: Both shutter speed (t_e) and inter-row delay (t_r) are required a priori in state-of-the-art RS-BMD [11] to fragment the motion trajectory for each image-row. Getting t_r from a camera requires processing of videos taken using the same camera setting (section 5.2 in [11]). Deriving both t_e and t_r *without* the meta-data and camera information further escalates the difficulty. In contrast, our method does *not* need any a priori timing information. Note that we estimate the value t_r/t_e for the RS prior in Eq. (9) solely from image intensities as discussed in section 4.4.

S2. Proof of Claim 2

Claim 2: *The prior which restricts drifting of TSFs between blocks (in Eq. (9)) is a convex function in* $\mathbf{w}$*, and can be represented as a norm of matrix vector multiplication,* i.e.*, as* $\|\mathbf{G}\mathbf{w}\|_2^2$*, with sparse* $\mathbf{G}$.

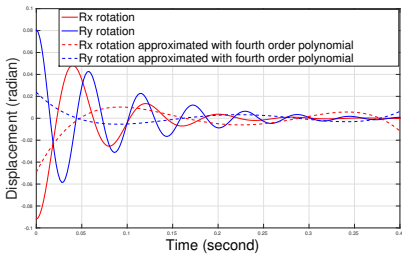


Figure S3. Vibrational ego-motion of a robotic system (using [4]). The polynomial model is *inadequate* to capture these trajectories.

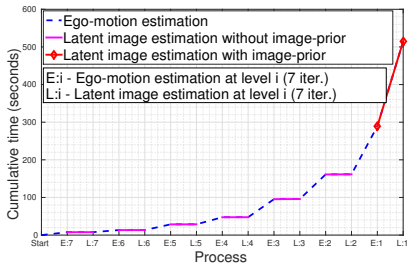


Figure S4. Cumulative time for different processes. Note the computational gains of the prior-less RS-EFF based image estimation.

To prove this, we draw from the following well-known properties of convex function [1] which are a linear function is always convex (prop. 1), composition of convex functions is always convex (prop. 2), and non-negative sum of convex functions is convex (prop. 3).

Proof: Considering n_b number of image blocks and each block-MDF $w_{\mathbf{i}}$ having length l, an individual additive component in our RS prior (in Eq. (9)) can be represented as $\|\Gamma\left(r_b(\mathbf{j}-\mathbf{i}+1)\right)\cdot\mathbf{S}_{(\mathbf{i},\mathbf{j})}\mathbf{w}\|_2^2$, where $\mathbf{S}_{(\mathbf{i},\mathbf{j})}$ is a matrix of dimension $l \times n_b \cdot l$, with all zeros except two scaled identity matrices of dimension $l \times l$ corresponding to $\mathbf{i}$th TSF (with scale 1) and $\mathbf{j}$th TSF (with scale -1). Therefore, the term $\{\Gamma\left(r_b(\mathbf{j}-\mathbf{i}+1)\right)\cdot\mathbf{S}_{(\mathbf{i},\mathbf{j})}\mathbf{w}\}$ is a linear function in $\mathbf{w}$. Since $\|\Gamma\left(r_b(\mathbf{j}-\mathbf{i}+1)\right)\cdot\mathbf{S}_{(\mathbf{i},\mathbf{j})}\mathbf{w}\|_2^2$ is a composite of squared L_2 norm (which is convex) of a linear function in $\mathbf{w}$, each additive component is convex (props. 1 and 2). Resultantly, the sum of all additive components in Eq. (9), *i.e.*, prior($\mathbf{w}$), is a convex function in $\mathbf{w}$ (prop. 3).

Also, prior($\mathbf{w}$) can be represented as $\|\mathbf{G}\mathbf{w}\|_2^2$, where matrix $\mathbf{G}$ is obtained by vertically concatenating matrices $\{\Gamma\left(r_b(\mathbf{j}-\mathbf{i}+1)\right)\cdot\mathbf{S}_{(\mathbf{i},\mathbf{j})}\}$ corresponding to the individual additive component in RS prior. Since $\mathbf{S}_{(\mathbf{i},\mathbf{j})}$ is a sparse matrix, $\mathbf{G}$ will also be sparse. Hence proved. ■

S3. Implementation Details

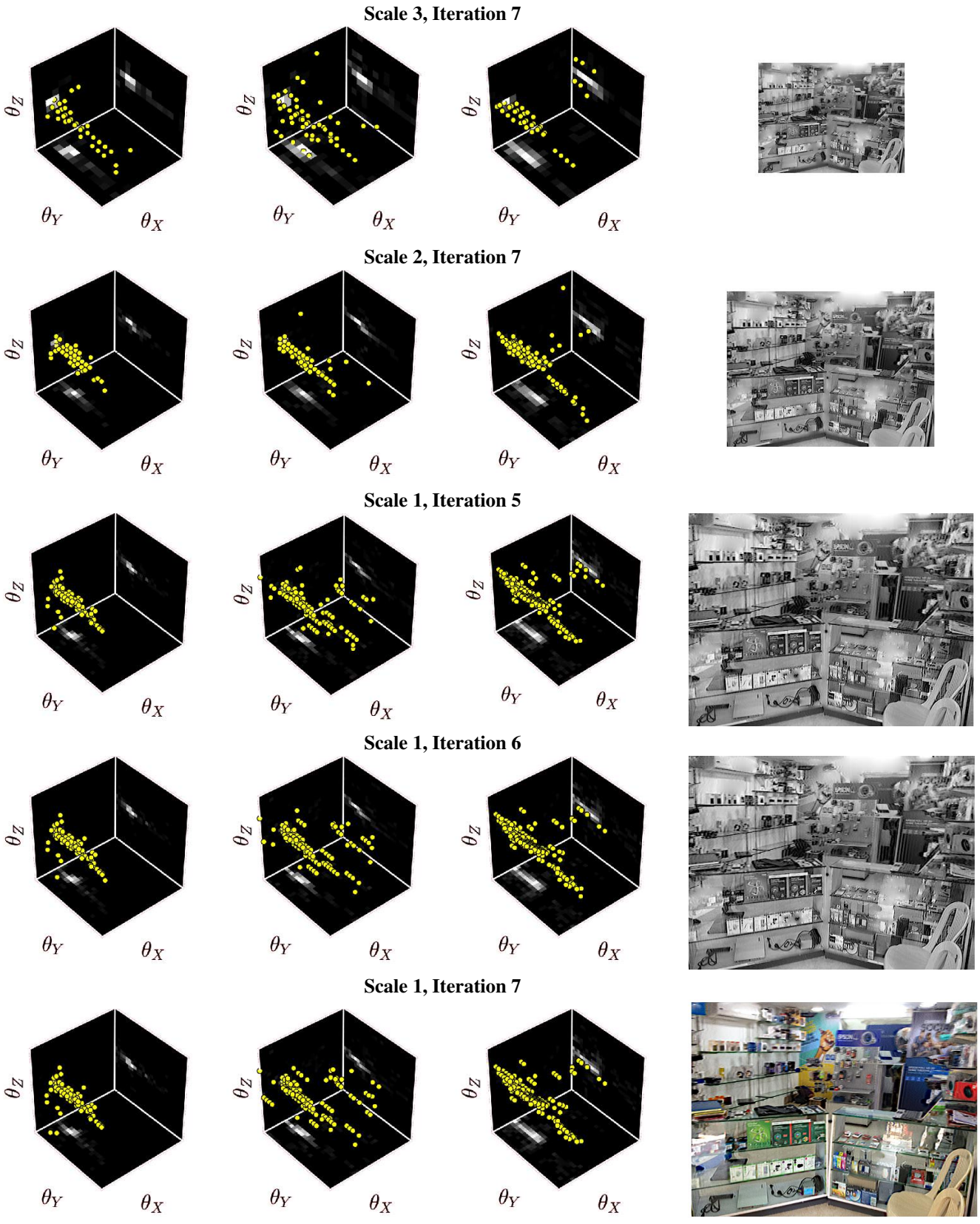
We implemented our algorithm in MATLAB. We empirically set 7 scales, each with 7 iterations, in our scale-space framework (section 4). The blurred image in the ith scale is formed by downscaling the input image by a factor of $(1/\sqrt{2})^{i-1}$. To start the alternative minimization, the coarsest scale MDFs are initialized with Kronecker delta. For ego-motion estimation (section 4.2), we consistently used the RS-prior regularization (α in Eq. (13)) in level i as 2^{7-i} (so that the RS prior can cope with the increasing image size, and thus the data fidelity magnitude $\|\mathbf{F}\mathbf{w}-\nabla\mathbf{B}\|_2^2$, in finer levels). We used the MDF regularization β' (in Eq. (13)) as 0.01. For latent image estimation (section 4.3), we used $R = 48$ such that each image-patch is square, and with 6 patches along the shorter dimension and 8 along the longer dimension. For the Richardson-Lucy deconvolution (employed in the last iteration of the finest level), we used a total number of 30 iterations. For the selection of block-size (section 4.4), we used an initial block-size r_0 as 145, and a downscaling factor of 2 (*i.e.*, $M_0 = M/2$ and $N_0 = N/2$).

Running time reported in Table 1 is obtained on the same system with an Intel Xeon processor with 32 GB memory. We found that for deblurring an 800×800 RGB image (of maximum blur-length of 30 pixels), our unoptimized MATLAB implementation took about 9 minutes. Fig. S4 provides a detailed break-up of the time taken for each estimation step. In fact, observe that a large fraction of the total time is utilized for latent image estimation in the *final* iteration which involves a costly image-prior (see section 4.3). This underscores the importance of our efficient prior-less estimation in the initial iterations derived from RS-EFF (Eq. (14)).

S4. Additional Evaluations

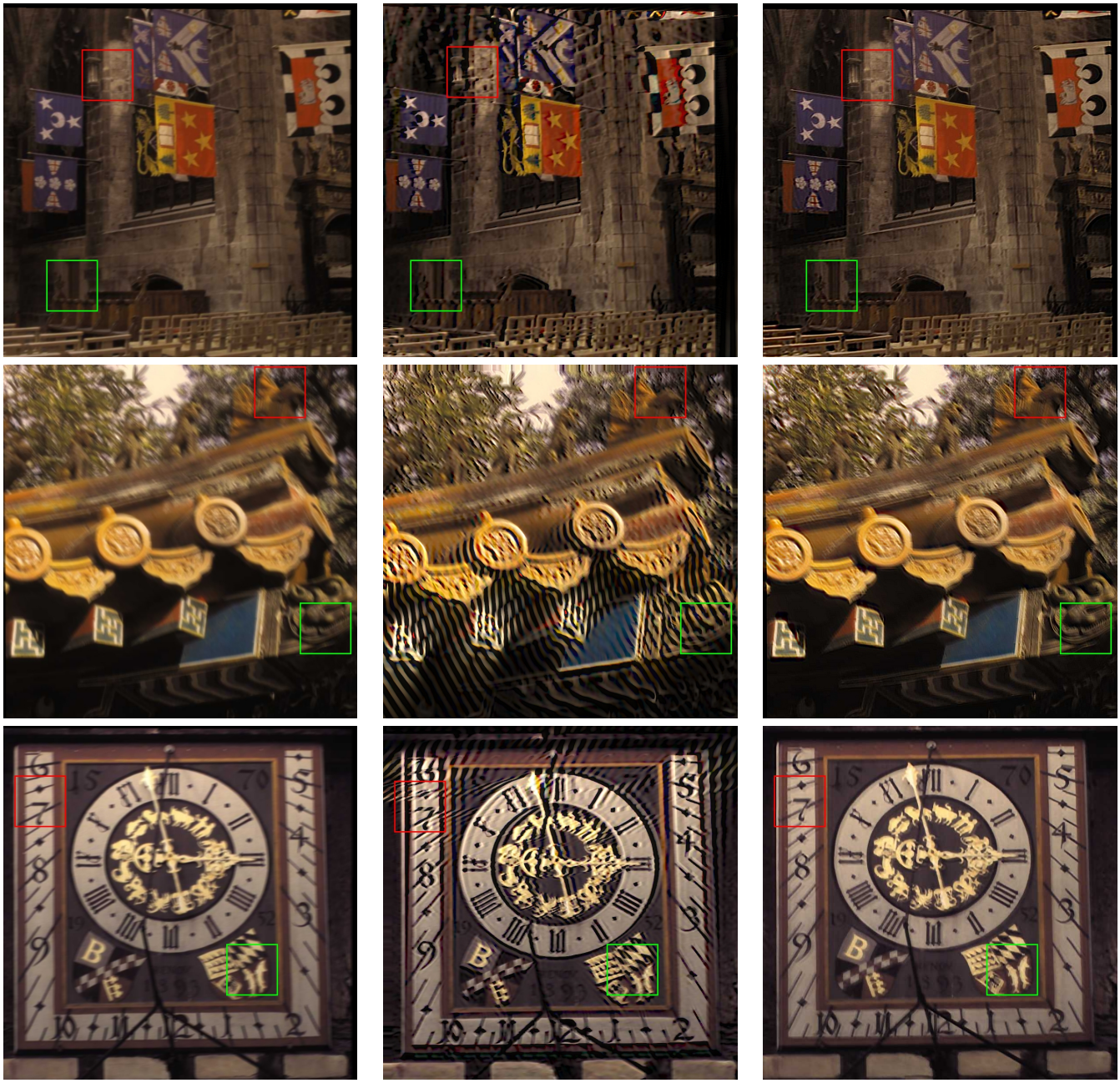
We provide in Fig. S5 iteration-by-iteration results to illustrate how the algorithm works. In Fig. S6, we give full images corresponding to the patches of synthetic experiment results provided in Figs. 5(a-i). In Figs. S7-S11, we give additional evaluations for the real RS-BMD examples provided in Figs. 7 & 8. These include SIV-BMD [2] and RS rectification followed with SIV-BMD [2] (as reported in [11]), and state-of-the-art CCD-BMD [9]. We also consider BMD *without* our RS prior to illustrate the ego-motion ambiguity in RS-BMD. For low-light case, we consider [5] that specifically addresses low-light BMD (albeit for CCD cameras). The codes for [5], [11] and [9] are downloaded from the author's website and executed using default parameters. Additional examples under different lighting condition and for wide-angle settings are given in Fig. S12.

For sake of completeness; we provide GS deblurring comparisons with state-of-the-art CCD-BMD methods of [9, 8, 10, 13, 12] and [3] in Figs. S13 & S14. We evaluated on the examples from the dataset of [7] and [9] using their reported results. The results show that our method works equally well for CCD cameras and importantly, *without* warranting any prior knowledge of the shutter.



(a) MDF 1 ($\mathbf{w_1}$) (b) MDF 2 ($\mathbf{w_2}$) (c) MDF 3 ($\mathbf{w_3}$) (d) Latent image ($\mathbf{L}$)

Figure S5. Iteration-by-iteration results of the alternative minimization of block-wise MDFs and latent image: (a-c) Estimated block-wise MDFs and (d) Estimated latent image. Notice the variation in block-wise MDFs, which depicts the characteristic of RS blur (as shown in Fig. 3). Also, observe the convergence of the block-wise MDFs through iteration 5 to 7 in the finest image scale (last three rows).



(a) Input (b) Su and Heidrich [11] (c) Ours

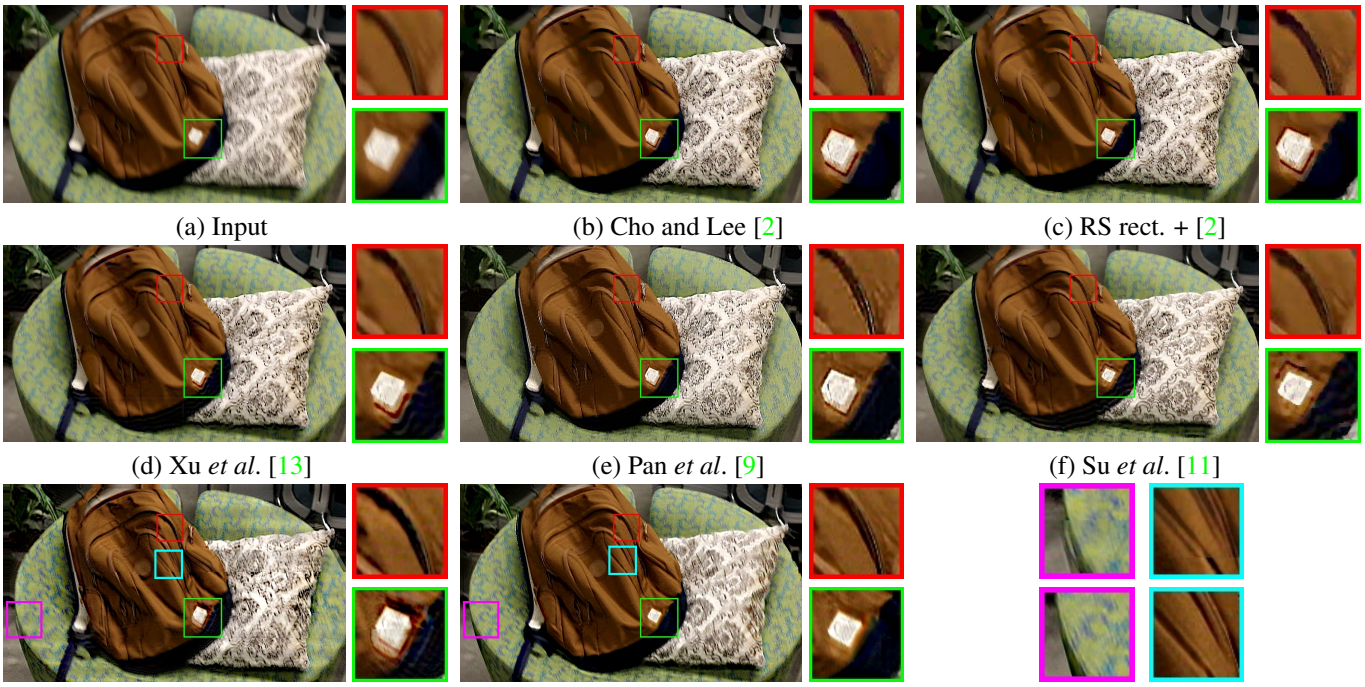
Figure S6. Full-sized images corresponding to the image patches given in Figs. 2(a-i): First row gives a case of wide-angle system (Figs. 2(a-c)), second row gives a case of vibratory motion (Figs. 2(d-f)), and third row gives a case of CCD-blur (Figs. 2(g-i)). (Best viewed on high-resolution display with zoom-in corresponding to an 800×800 image size.)

(a) Input (b) Cho and Lee [2] (c) RS rect. + [2]

(d) Xu *et al*. [13] (e) Pan *et al*. [9] (f) Su *et al*. [11]

(g) Ours *without* RS prior (h) Ours with RS prior (i) Patches of (g) in top, and (h) in bottom

Figure S7. Detailed comparisons for RS narrow-angle example in dataset [11] (Fig. 7-top-row). Note the effect of incoherent combination due to the block shift-ambiguity (section 3, claim 1) in (i)-first row, which is successfully suppressed by our RS prior ((i)-second row).

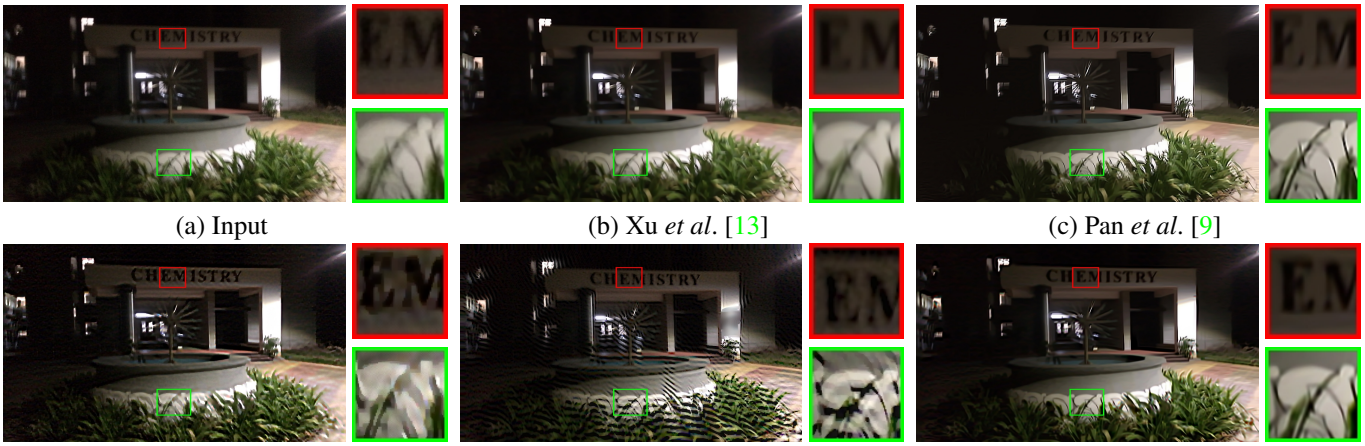


(a) Input (b) Cho and Lee [2] (c) RS rect. + [2]

(d) Xu *et al*. [13] (e) Pan *et al*. [9] (f) Su *et al*. [11]

(g) Ours *without* RS prior (h) Ours with RS prior (i) Patches of (g) in top, and (h) in bottom

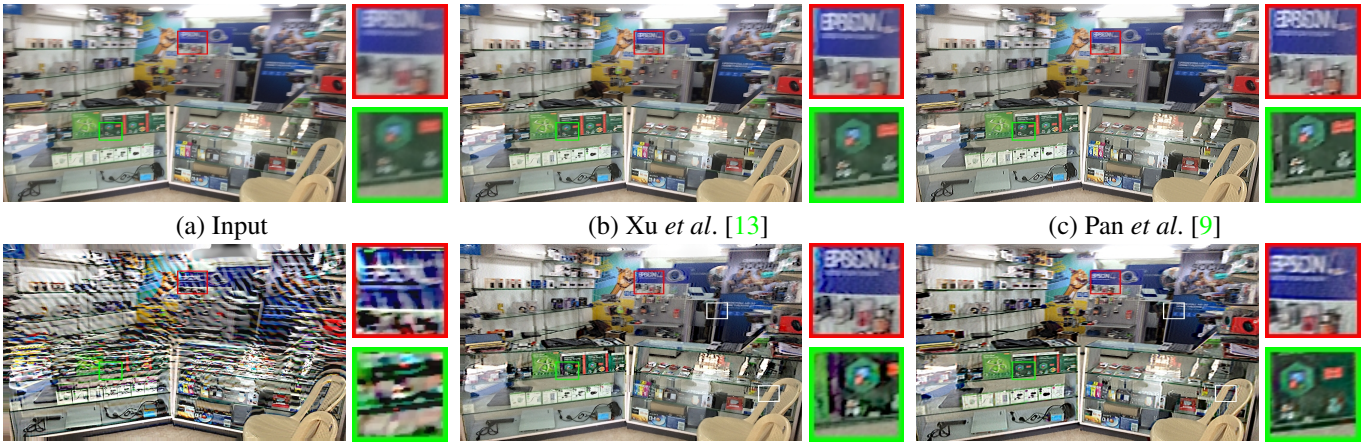
Figure S8. Detailed comparisons for RS narrow-angle example in dataset [11] (Fig. 7-second-row). Our method recovers finer details (see bag-zipper in patch 1), and deblur with negligible ringing artefacts (see bag-badge in patch 2), as compared to competing methods.



(a) Input (b) Xu *et al*. [13] (c) Pan *et al*. [9]

(d) Hu *et al*. [5] (e) Su *et al*. [11] (f) Ours

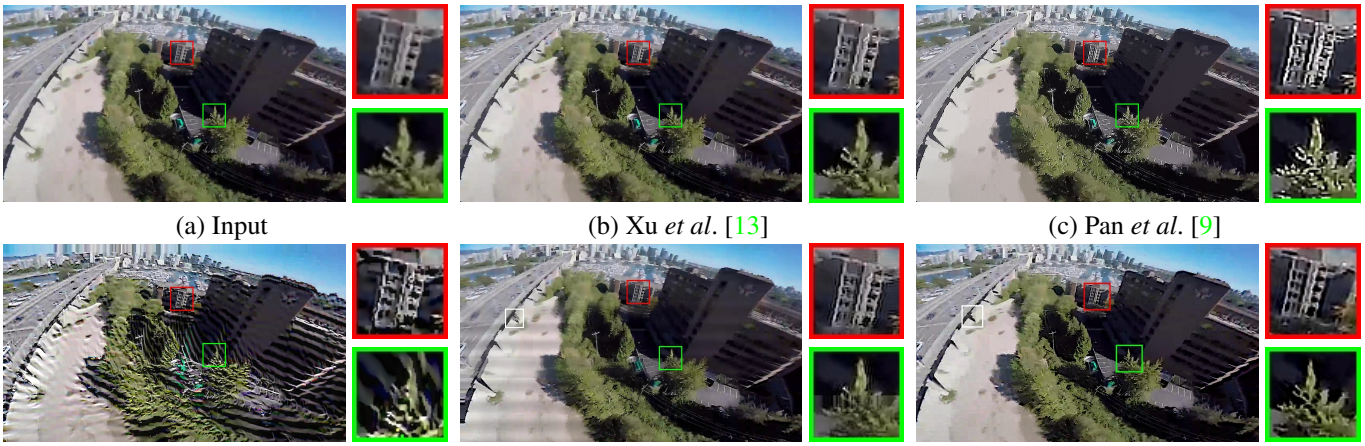
Figure S9. Detailed comparisons for RS wide-angle example (Fig. 8-first row). In contrast to competing methods, our method models the RS ego-motion better (observe the residual blur in the letters, and the repeated occurrence of the longest grass leaf in (c)).



(a) Input (b) Xu *et al*. [13] (c) Pan *et al*. [9]

(d) Su *et al*. [11] (e) Ours *without* RS prior (f) Ours

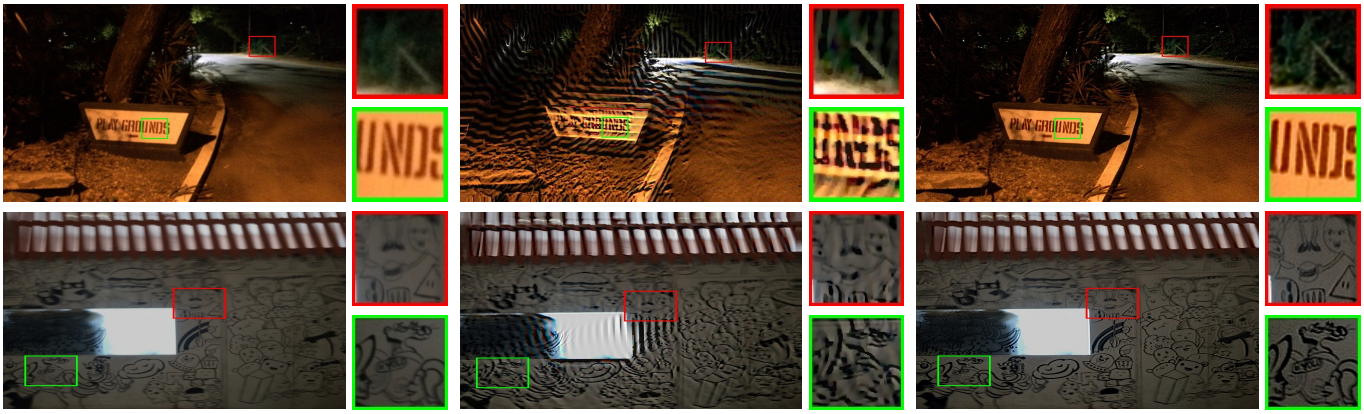
Figure S10. Comparisons for RS wide-angle case (Fig. 8-second row). White boxes in images (e) and (f) show the effect of RS prior.



(a) Input (b) Xu *et al*. [13] (c) Pan *et al*. [9]

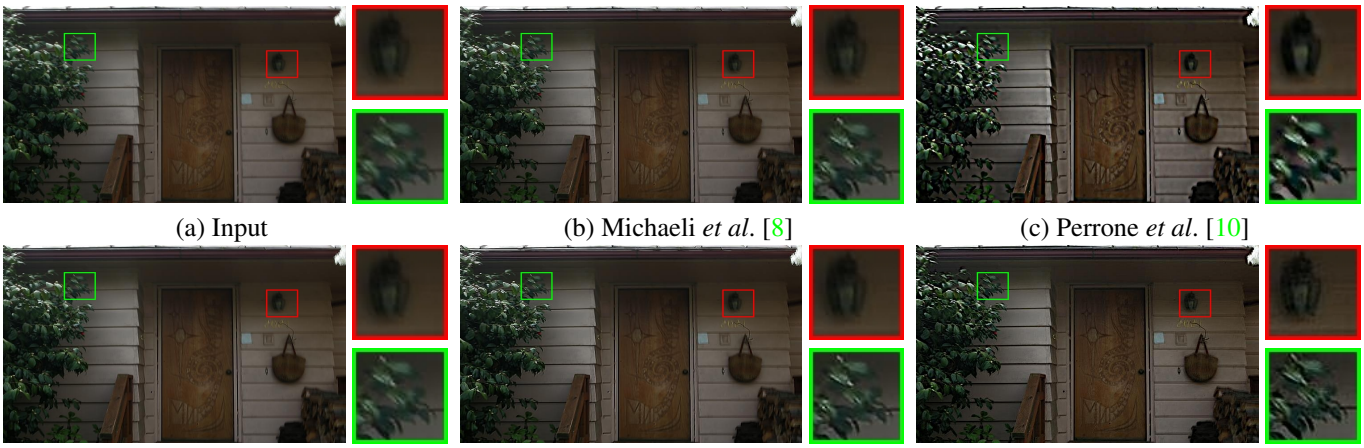
(d) Su *et al*. [11] (e) Ours *without* RS prior (f) Ours

Figure S11. Comparisons for RS wide-angle example (Fig. 8-third row). White box in images (e) and (f) shows the effect of RS prior.



(a) Input (b) Su *et al*. [11] (c) Ours

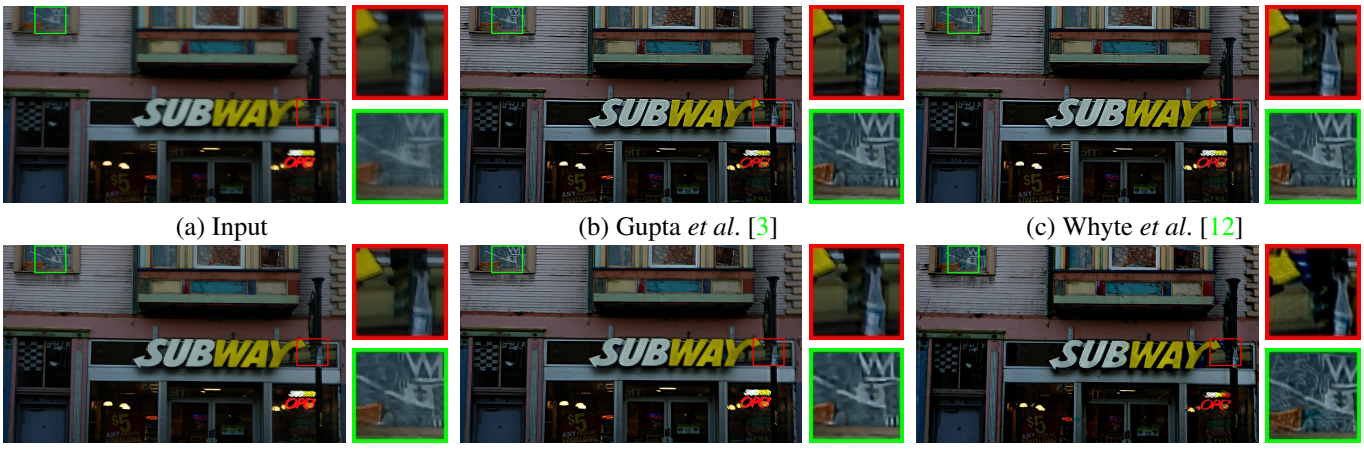
Figure S12. Additional RS comparisons with state-of-the-art RS-BMD method [11] under different lighting conditions and for wide-angle settings. Note the inefficacy of the competing method in dealing with wide angle systems.



(a) Input (b) Michaeli *et al*. [8] (c) Perrone *et al*. [10]

(d) Xu *et al*. [13] (e) Pan *et al*. [9] (f) Ours

Figure S13. Comparisons for CCD blur example in dataset [7]. Our result is comparable with [8, 10, 13] and [9].



(a) Input (b) Gupta *et al*. [3] (c) Whyte *et al*. [12]

(d) Xu *et al*. [13] (e) Pan *et al*. [9] (f) Ours

Figure S14. Comparisons for CCD blur example in dataset [9]. Our result is comparable with [3, 12, 13] and [9].

References

[1] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 2

[2] S. Cho and S. Lee. Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, page 145. ACM, 2009. 2, 5

[3] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *European Conference on Computer Vision (ECCV)*, pages 171–184. Springer, 2010. 2, 7

[4] M. R. Hatch. *Vibration simulation using MATLAB and ANSYS*. CRC Press, 2000. 1, 2

[5] Z. Hu, S. Cho, J. Wang, and M.-H. Yang. Deblurring low-light images with light streaks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3382–3389, 2014. 2, 6

[6] 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] 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. 2, 7

[8] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *European Conference on Computer Vision (ECCV)*, pages 783–798. Springer, 2014. 2, 7

[9] 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. 2, 5, 6, 7

[10] D. Perrone and P. Favaro. Total variation blind deconvolution: The devil is in the details. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2909–2916. IEEE, 2014. 2, 7

[11] 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, 2, 4, 5, 6, 7

[12] 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. 2, 7

[13] 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. 2, 5, 6, 7