Rolling Shutter Motion Deblurring

Shuochen Su[1,2], Wolfgang Heidrich[2,1]
[1]University of British Columbia. [2]KAUST.

Although motion blur and rolling shutter deformations are closely coupled artifacts in images taken with CMOS image sensors, the two phenomena have so far mostly been treated separately, with deblurring algorithms being unable to handle rolling shutter wobble, and rolling shutter algorithms being incapable of dealing with motion blur.

We propose an approach that delivers sharp and undistorted output given a single rolling shutter motion blurred image. The key to achieving this is a global modeling of the camera motion trajectory, which enables each scanline of the image to be deblurred with the corresponding motion segment (Fig. 1). Provided that the underlying camera pose at time t is known as $\mathbf{p}(t)$, we first write the image formation model for each row $\mathbf{b}_i$ in a RSMB image $\mathbf{B} = (\mathbf{b}_0^T, \ldots, \mathbf{b}_M^T)^T$ as

$$\mathbf{b}_i = \frac{1}{t_e}\int_{i\cdot t_r}^{i\cdot t_r + t_e} \mathbf{l}_i^{\mathbf{p}(t)} dt + \mathbf{n}_i \tag{1}$$

where $\mathbf{l}_i^{\mathbf{p}(t)}$ and $\mathbf{n}_i$ represent the i-th row in the transformed latent image and the noise image respectively.

Eq. (1) can be expressed in discrete matrix-vector form after assuming a finite number of time samples during the exposure of each row, so that $b_{ij} \in \mathbf{b}_i$ can be exactly determined at any pixel location $\mathbf{x} = (i, j)^T$ in $\mathbf{B}$ as

$$b_{ij} = \frac{1}{|\mathbb{T}_i|}\sum_{t\in\mathbb{T}_i} \Gamma_{\mathbf{L}}(\mathbf{w}(\mathbf{x};\mathbf{p}(t))) + n_{ij}, \tag{2}$$

where $\mathbb{T}_i = \{i\cdot t_r + \frac{j}{K}t_e\}_{j=0\ldots K}$ is a set of uniformly spaced time samples in the exposure window of row i, $\Gamma_{\mathbf{L}}(\cdot)$ is the function that bi-linearly interpolates the intensity at some sub-pixel position in $\mathbf{L}$, and $\mathbf{w}(\cdot)$ is a warping function that maps positions $\mathbf{x}$ from the camera frame back to the reference frame of the latent image $\mathbf{L}$ according to the current camera pose $\mathbf{p}$ [1]. Based on an analysis over the publicly available camera motion trajectories from Kohler *et al.* [3] we also decide to fit polynomial functions to the pose trajectories: $\mathbf{p}(t) = \mathbf{t}\theta$, where $\mathbf{t} = (t^P, \ldots t^0)$, $t \in [0, Mt_r + t_e]$, and θ is a coefficient matrix. From here, a sparse convolution matrix $\mathbf{K}$ can be created and Eq. (1) can be rewritten as

$$\mathbf{b} = \mathbf{K}(\theta)\mathbf{l} + \mathbf{n}, \tag{3}$$

where $\mathbf{b}$, $\mathbf{l}$ and $\mathbf{n}$ respectively are the RSMB input, latent image, and noise, all in vector form.

Having the forward RSMB model defined in the form of a camera motion model, the latent image $\mathbf{l}$ can be recovered from $\mathbf{b}$ by solving an inverse problem, where the objective function is given by

$$\min_{\mathbf{l},\theta} \frac{1}{2}\|\mathbf{b} - \mathbf{K}(\theta)\mathbf{l}\|_2^2 + \mu\|\nabla\mathbf{l}\|_1, \tag{4}$$

Similar to conventional blind deblurring algorithms, we update $\mathbf{l}$ and θ in an alternating fashion. We initialize μ with a relative large value μ_0, thus in the early iterations only the most salient structure in $\mathbf{l}$ will be preserved which will guide the refinement of kernel coefficients θ, given that θ estimates are not yet accurate. As the optimization progresses, we decrease μ by a factor of τ after each iteration to preserve more details in $\mathbf{l}$.

The objective for updating θ is given by

$$\theta^{k+1} = \arg\min_{\theta} \sum_{i=0}^{M}\sum_{j=0}^{N} r_{ij}(\theta)^2$$
$$\text{where } r_{ij}(\theta) = b_{ij} - \frac{1}{|\mathbb{T}_i|}\sum_{t\in\mathbb{T}_i}\Gamma_{\mathbf{L}^k}(\mathbf{w}(\mathbf{x};\mathbf{t}\theta)) \tag{5}$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

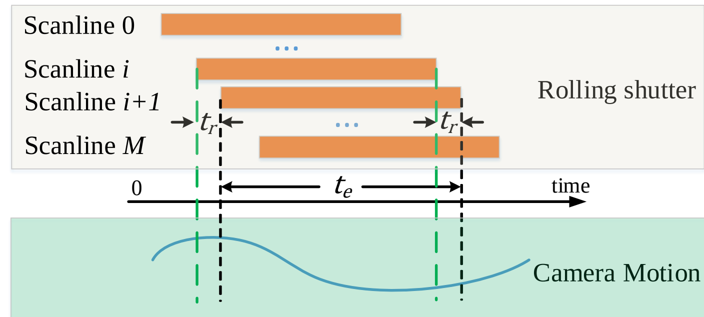

Figure 1: Different scanlines integrate over a slightly different segment of the motion trajectory, resulting in a shift-variant kernel.

Figure 2: Insets of real RSMB images (top) and their deblurred results from our method (bottom).

Here $\mathbf{L}^k$ is the latent image estimated from the previous iteration k. Solving Eq. (5) is a non-linear optimization task since pixel values in $\mathbf{L}$ are, in general, non-linear in θ. Motivated by image registration, *i.e.* the Lucas-Kanade algorithm [1], we adopt Gauss-Newton method for this non-linear least square problem, where the initial trajectory estimates are fitted by solving blind deconvolution problems for blocks of several scanlines.

Fixing θ^k, we solve the latent image update subproblem

$$\mathbf{l}^{k+1} = \arg\min_{\mathbf{l}} \frac{1}{2}\|\mathbf{K}(\theta^k)\mathbf{l} - \mathbf{b}\|_2^2 + \mu\|\nabla\mathbf{l}\|_1. \tag{6}$$

where the alternating direction method of multipliers (ADMM) algorithm [2] is used to address the presence of L1 norm in the objective.

We perform a series of experiments on both synthetic and real (Fig. 2) RSMB images, and conduct quantitative and qualitative comparison with conventional blind deblurring work, including the strategy of first rectifying the rolling shutter wobble from videos containing the specific frames before applying conventional blind deblurring algorithms. See the main paper and supplemental material for details.

[1] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56(3):221–255, 2004.

[2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[3] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *Computer Vision–ECCV 2012*, pages 27–40. Springer, 2012.