ROLLING SHUTTER SUPER-RESOLUTION IN BURST MODE

Vijay Rengarajan, Abhijith Punnappurath*, A.N. Rajagopalan*, Gunasekaran Seetharaman*[†]

[*]Indian Institute of Technology Madras [†]US Naval Research Laboratory

ABSTRACT

Capturing multiple images using the burst mode of handheld cameras can be a boon to obtain a high resolution (HR) image by exploiting the subpixel motion among the captured images arising from handshake. However, the caveat with mobile phone cameras is that they produce rolling shutter (RS) distortions that must be accounted for in the super-resolution process. We propose a method in which we obtain an RS-free HR image using HR camera trajectory estimated by leveraging the intra- and inter-frame continuity of the camera motion. Experimental evaluations demonstrate that our approach can effectively recover a super-resolved image free from RS artifacts.

Index Terms— Burst mode, rolling shutter, super-resolution

1. INTRODUCTION

The burst mode feature available in most cameras today allows the user to take a sequence of shots in rapid succession with a single click. This can be leveraged to obtain a higher resolution image from the captured low resolution (LR) images using super-resolution (SR) techniques [1, 2, 3, 4, 5]. Classical multi-image SR algorithms assume that the motion between images is global since they are designed for CCD cameras where all sensors in the sensor plane are exposed at once during the exposure duration of an image. On the contrary, this is not true for CMOS sensors that are commonly employed in mobile phones, in which each row of sensors has its own unique exposure duration, and the gathered data is read out sequentially. The problem with this row-wise acquisition scheme is that the motion of either the camera or the scene during exposure will lead to geometric distortions in the image since each row is captured at a different time. This distortion of the acquired image is called the rolling shutter (RS) effect, which results from each row of sensors observing a different warp of the scene. Hence, one must obtain the camera trajectory throughout the row exposures to perform super-resolution.

There are a plethora of works that estimate the row-wise camera motion from the RS distorted images themselves for various applications, including video stabilization [6, 7, 8, 9], change detection [10], deblurring [11], and structure from motion [12, 13]. One could also obtain the camera motion using the built-in inertial sensors [14, 15, 16], but the flip side is their low acquisition rate and the need for their synchronization with image exposure time, especially since super-resolution needs an accurate camera motion estimate.

The only work in the literature to attempt SR for RS images is the recent work of Punnappurath et al. [17]. They propose an RS-SR observation model and a unified framework to obtain an RS-free HR image from distorted LR images captured using a CMOS camera. However, they assume that one of the images is free from RS effect, and use this reference image to estimate the row-wise motion of the other images. This assumption severely limits the practical applicability of their method because all images will typically have RS effect when captured using a hand-held camera. To simplify the model, they also assume that a block of rows (with the block size being equal to the super-resolution factor) at HR experience the same warp. Strictly speaking, this is not true because *each* row of the HR image can have its own motion.

In this paper, we present an approach to recover the undistorted HR image even when all the input images are RS-affected. The LR images are captured in the burst mode using a hand-held RS camera allowing us to represent the continuous trajectory using a sparse set of camera poses. The point correspondences needed to estimate the camera path are obtained using SIFT [18] feature matches between the images. Since the HR camera motion and the HR image are both unknown, we propose an alternating minimization (AM) strategy to solve for the two variables. To initialize the algorithm, an approximate HR trajectory is constructed using an estimated LR motion trajectory that corrects the RS effect in all LR frames. In contrast to Punnappurath et al. [17] where a block of rows at HR is enforced to have the same motion, our formulation allows each HR row to have its own motion.

Contributions:

- To the best of our knowledge, this is the first attempt at joint rectification and super-resolution from multiple LR images that are *all* RS-affected.
- Unlike the current state-of-the-art work [17] where a warp at HR is associated with a block of rows, we strictly adhere to the observation model and allow each row of the HR image to have its own motion.

2. ROLLING SHUTTER SUPER-RESOLUTION

Let $\{\mathbf{g}_k\}_{k=1}^{K}$ represent the set of K LR images acquired in succession using the burst mode of the camera. Each $\mathbf{g}_k$ has M rows and N columns, and all $\mathbf{g}_k$s are RS-affected due to camera motion. Let $\mathbf{f}$ denote the underlying clean HR image having mM rows and mN columns, where m is the super-resolution factor. Let the camera motion during the burst capture be denoted by a vector of 6D camera poses $\mathbf{p}$. Each element of $\mathbf{p}$ is associated with one row of $\mathbf{f}$. The number of pose elements in $\mathbf{p}$ is given by mKM corresponding to mM rows in $\mathbf{f}$ and K image captures. Thus, the entire pose vector $\mathbf{p}$ can be viewed as the concatenation of K smaller pose vectors corresponding to K LR images, and each of these K pose vectors is denoted by $\mathbf{p}_k$. During the image formation process, $\mathbf{f}$ is affected by the HR-RS camera motion $\mathbf{p}$ during K image exposures, and each $\mathbf{p}_k$ produces one LR-RS image $\mathbf{g}_k$ after downsampling. Let the time delay between any two successive image exposures in burst mode be t_b, which can be equivalently represented as an exposure of a fixed number of blank rows, n_b. Thus, $\mathbf{p}$ contains the discretized camera poses for each image row of the continuous camera motion with discontinuities during the exposure of blank rows (i.e. during the idle time after each image exposure). The relationship between the HR image and the LR images is given by

$$\mathbf{g}_k = \mathbf{D}_m \mathbf{R}_{\mathbf{p}_k} \mathbf{f}, \quad \text{for } k = 1, \ldots, K, \tag{1}$$

where $\mathbf{D}_m \in \mathbb{R}^{MN \times m^2 MN}$ is the downsampling operator and $\mathbf{R}_{\mathbf{p}_k}$ is the matrix equivalent of the vector of poses $\mathbf{p}_k$ that warps the image $\mathbf{f}$ to produce an RS image. Given the LR-RS images $\{\mathbf{g}_k\}_{k=1}^{K}$, our goal is to estimate both the underlying HR camera motion, $\mathbf{p}$, and the HR image, $\mathbf{f}$.

We follow an alternating minimization (AM) framework in which the HR camera motion and the HR clean image are estimated in an iterative manner. In the camera motion estimation step, the goal is to estimate the camera motion experienced by the HR image which produces the observed LR-RS images after downsampling. In the image estimation step, an HR image is estimated given the HR camera motion and the LR-RS images. Since all LR images are RS affected, an RS rectification step is required, and hence, we estimate an RS-corrective LR trajectory using which an initial HR trajectory is obtained to kick-start the HR image estimation step. We then alternate between the two AM steps refining both the HR motion and the HR image in every iteration.

2.1. Camera Motion Estimation

Given an estimated clean HR image $\widehat{\mathbf{f}}$ and the observed LR-RS images $\{\mathbf{g}_k\}$, we estimate the underlying HR-RS camera motion following a feature-point based approach. For each of the HR-LR pairs $\widehat{\mathbf{f}}$-$\mathbf{g}_k$, we detect SIFT feature point matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. We then estimate the HR camera motion $\mathbf{p}$ that warps $\mathbf{x}_i$ to $\mathbf{x}'_i$ after downsampling.

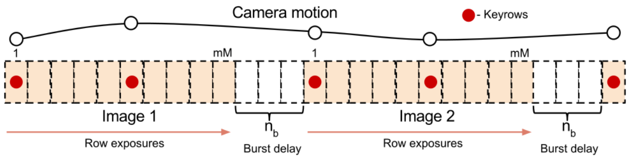

Fig. 1. Motion model with two keyrows per image ($q = 2$) for two image exposures ($K = 2$). The total number of keyrows throughout the total exposure duration is thus 5.

Since the number of unknown poses in $\mathbf{p}$ is too high (mKM), we use a key-row interpolation approach that treats only the camera poses at certain rows as unknowns. These unknown camera poses correspond to q equally spaced key-rows in $\widehat{\mathbf{f}}$ for each $\widehat{\mathbf{f}}$-$\mathbf{g}_k$ pair, where $q \ll mM$. The HR sub-trajectory $\mathbf{p}_k$ corresponding to $\mathbf{g}_k$ is determined by the camera poses located in the key-rows, $\{1 \cup r.mM/q : r = 1, 2, \ldots, q-1\}$ of the HR image. The last part of the HR trajectory corresponding to $\mathbf{g}_K$ has an extra key-row after its last row mM. Hence, the number of unknown poses in $\mathbf{p}$ reduces to $qK + 1$ corresponding to the total number of key-rows. Fig. 1 shows the key-row locations for two image exposures having two key-rows per exposure, i.e. for $K = 2$ and $q = 2$. Any other camera pose corresponding to a row situated between two key-rows is modelled as an interpolation of these two key-row poses using linear interpolation for translations and spherical linear interpolation for rotations [9, 19]. The pose interpolation for rows situated between the last key-row of an image and the first key-row of the next image needs the value of the number of blank rows n_b corresponding to the burst delay t_b. Hence, we additionally estimate n_b as well jointly with key-row poses. Though the camera trajectory is defined using the camera poses situated only at key-rows, each row of the HR image can have a unique camera pose in our model due to interpolation. We observed that this model is simple and effective for modelling camera handshake.

To estimate $\mathbf{p}$ and n_b, we employ a cost based on the point correspondences in the HR and LR images. The camera motion when acted upon $\mathbf{x}_i$ in the HR image results in an HR-RS point, which when downsampled by the factor m results in the LR-RS point $\mathbf{x}'_i$. We accumulate the resultant error in this operation for all point correspondences and form the following optimization problem:

$$\widehat{\mathbf{p}} = \arg\min_{\mathbf{p}, n_b} \left\{ \sum_{k=1}^{K} \sum_{\substack{\mathbf{x}_i \in \mathbf{f} \\ \mathbf{y}_i \in \mathbf{g}_k}} \left\| \frac{1}{m} \mathbf{p}_k(\mathbf{x}_i) - \mathbf{x}'_i \right\|_2^2 \right\} \tag{2}$$

where $\mathbf{p}_k(\mathbf{x}_i)$ denotes the resultant point obtained when $\mathbf{x}_i$ is warped using $\mathbf{p}_k$. The operation $\mathbf{p}_k(\mathbf{x}_i)$ warps $\mathbf{x}_i$ using a camera pose determined by interpolation of key-row poses based on the row coordinate of the LR-RS coordinate $\mathbf{x}'_i$ multiplied by m. The nonlinear least squares minimization (2) is solved using `lsqnonlin` function in MATLAB. Our ap-

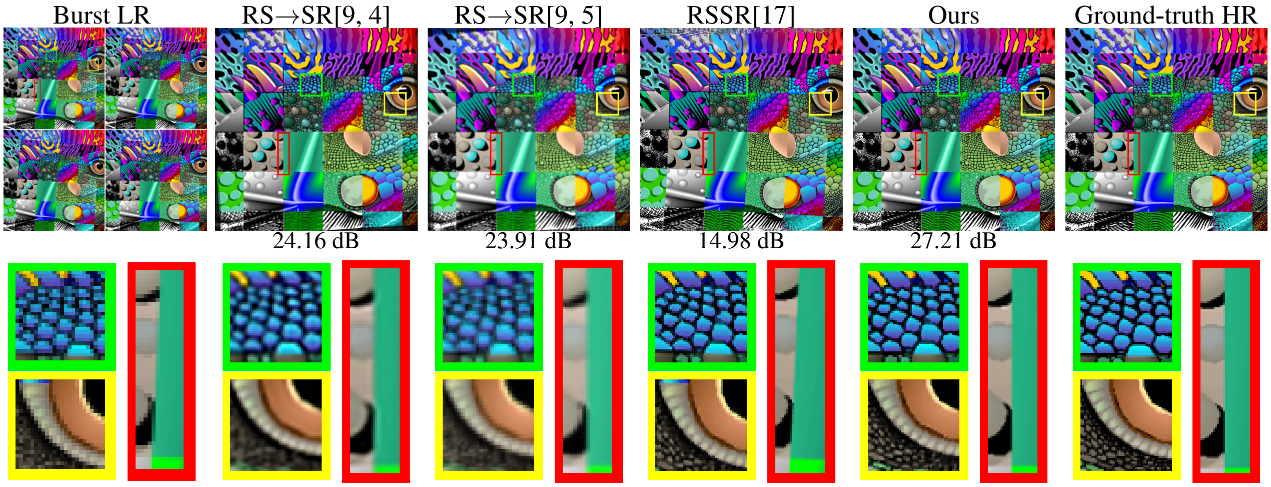

Fig. 2. *Synthetic experiment*: Input LR images, super-resolved outputs from different methods, and ground-truth HR image.

proach to register RS images is feature-based unlike the existing RS super-resolution work of [17] which is intensity-based, and hence our motion estimation algorithm is faster.

2.2. Image Estimation

Given an HR motion estimate $\widehat{\mathbf{p}}$, we create the RS motion matrices $\{\mathbf{R}_{\widehat{\mathbf{p}}_k}\}$ from each $\widehat{\mathbf{p}}_k$ and solve the following optimization problem to obtain a clean HR image $\widehat{\mathbf{f}}$ that relates the given LR-RS images $\{\mathbf{g}_k\}$ via the HR motion $\widehat{\mathbf{p}}$.

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f}} \left\{ \sum_{k=1}^{K} \|\mathbf{g}_k - \mathbf{D}_m \mathbf{R}_{\widehat{\mathbf{p}}_k}\mathbf{f}\|_2^2 + \alpha \mathbf{f}^T\mathbf{L}\mathbf{f} \right\} \qquad (3)$$

where $\mathbf{L}$ is the discrete form of the total variation prior [20] which favors local smoothness while also preserving edges, and α is a non-negative weight. We used a fixed value of 1500 for α in all our examples based on visual assessment. We solve (3) using the method of conjugate gradients [20, 17] (`pcg` function in MATLAB).

2.3. Initial Motion Estimate

The assumption of the presence of an RS-free LR image allowed the work of [17] to jumpstart their algorithm by upsampling. In our case, it is essential to estimate the LR corrective camera motion that would remove the RS effect from all the LR images, since none of them is RS-free. We bootstrap the AM algorithm with an initial HR motion estimate obtained by estimating the LR corrective camera motion from the given LR-RS images using a feature-point approach similar to that in Sec. 2.1.

We first obtain the point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ *between* each consecutive LR image pair. We then seek an LR motion $\widehat{\mathbf{p}}_{LR}$ spanning the whole K exposure period that would correct the RS effect in *all* LR images to a virtual clean reference LR image respecting the relationship between correspondence points of all LR image pairs by solving the following optimization problem:

$$\min_{\substack{\mathbf{p}_{LR}\\ n_{b,LR}}} \left\{ \sum_{k=1}^{K-1} \sum_{\substack{\mathbf{x}_i\in\mathbf{g}_k\\ \mathbf{y}_i\in\mathbf{g}_{k+1}}} \left\| \mathbf{p}^{-1}_{LR,k}(\mathbf{x}_i) - \mathbf{p}^{-1}_{LR,k+1}(\mathbf{x}'_i)\right\|_2^2 \right\} \qquad (4)$$

Here $\mathbf{p}^{-1}_{LR,k}(\mathbf{x}_i)$ denotes the RS correction of $\mathbf{x}_i$ by warping it using the inverse of the camera motion $\mathbf{p}_{LR,k}$, and $n_{b,LR}$ is the number of blank rows corresponding to the LR exposure. After solving for the LR motion estimate $\widehat{\mathbf{p}}_{LR}$ using (4), the initial HR motion estimate $\widehat{\mathbf{p}}$ is obtained from $\widehat{\mathbf{p}}_{LR}$ through doubling of translations and replication of motion of each row in LR to m consecutive rows of HR. Note that this replication is done only for initialization. The initial HR estimate $\widehat{\mathbf{p}}$ is provided as input to (3) to start the AM framework with HR image estimation.

3. EXPERIMENTS

We show two RS-SR examples in the paper; the first is a synthetic experiment where we quantitatively analyze the results with the ground-truth, and the second is a real experiment where the images are captured in burst mode using a handheld mobile phone. We compare our results with (i) the existing state-of-the-art RSSR method of [17] which requires an RS-free reference image, and (ii) a serial framework of RS-rectification [9] followed by multi-image classical SR techniques [4, 5] to tackle RS and SR independently using their respective state-of-the-art methods (denoted by RS→SR).

Synthetic experiment The burst mode capture is simulated by synthesizing a continuous camera motion to create four RS images at HR and then downsample them to create four LR-RS images. The motion applied in inplane rotations which are commonplace in handshake, and they create visible

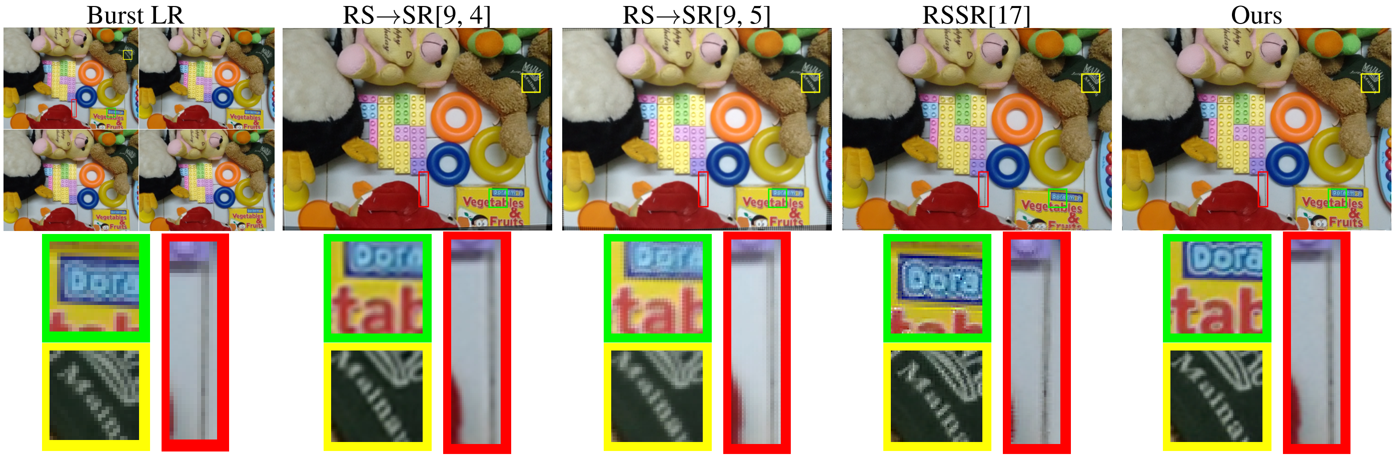

Fig. 4. *Real experiment*: Input LR images and super-resolved outputs from different methods.

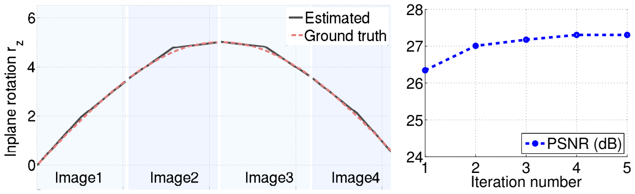

Fig. 3. *Synthetic example* Left: Estimated HR motion. Right: PSNR over AM iterations.

RS effect even for small rotation values. The size of the HR image in pixels is 512×512 and the number of blank rows simulated between two image exposures is 20. The down-sampling factor applied is two and the super-resolution factor m is also the same. Fig. 2 shows the LR images with size in pixels as 256×256. The zoomed-in patches from the first LR-RS image is shown in the second row. The square patches show the lower resolution and the long patch shows a slanted edge due to the curvature formed by the RS effect. We also show the ground-truth HR image in the sixth (last) column of Fig. 2 for comparison.

For the sequential RS→SR methods, we first rectify the LR images using the RS rectification method of [9], and then perform multi-image SR using [4] and [5]. The second and third columns of Fig. 2 show the output images and patches. Even though the sequential approach corrects the RS effect, its SR performance is not very good as can be observed from the PSNR values (24.16 dB and 23.91 dB) due to the lack of refinement of camera motion iteratively. To compare with the joint RSSR method of [17], we consider the first LR-RS image as the reference image. We then perform joint motion estimation and super-resolution using [17]. The visual quality of the SR image is good, but the RS effect is not removed. This can be observed from the retained slant of the edge in the long patch. Hence, it provides a low PSNR of 14.98 dB. Our method takes all LR-RS images as inputs and performs joint RS rectification and SR. Our RS-free super-resolved HR image is shown in the fifth column of Fig. 2 and the corresponding patches show the absence of RS artifacts as well as an increase in visual quality with a PSNR of 27.21 dB.

We also show our estimated HR camera trajectory along with the ground truth in Fig. 3. Our estimated path closely follows the ground truth during all image exposures. We used $q = 2$ in this experiment. The white empty strips between two image exposures indicate the burst delay, and our method adapts to it and estimates the camera motion correctly. We also show the output PSNR after each AM iteration in Fig. 3. There is a sharp increase in PSNR after the first iteration, and the AM framework converges after five iterations.

Real experiment We capture a set of images in burst mode using a Moto G2 mobile phone with handshake. Fig. 4 shows the input LR images and the SR output images obtained using different methods. Our method clearly outperforms the comparison methods. It has superior visual quality over sequential RS→SR methods of [9, 4] and [9, 5] as can be observed from the clearer text in square patches. It corrects the RS effect too, unlike the existing joint RSSR method of [17] as can be observed from the line in the long patch.

We have provided additional real examples in the supplementary PDF.

4. CONCLUSIONS

In this paper, we addressed the challenging problem of super-resolution from multiple RS-affected images. Unlike the current state-of-the-art which assumes a clean reference image, our method is capable of recovering an undistorted and super-resolved image even when *all* the input images are RS-affected. Experiments reveal that our method significantly advances the state-of-the-art. As future work, we would like to extend our formulation to model the effects of both blur and rolling shutter in the LR frames.

5. REFERENCES

[1] David Peter Capel, "Image mosaicing and superresolution," 2004.

[2] Sina Farsiu, M. Dirk Robinson, Michael Elad, and Peyman Milanfar, "Fast and robust multiframe super resolution.," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[3] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Variational bayesian super resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 984–999, April 2011.

[4] Alfonso Snchez-Beato, "Coordinate-descent super-resolution and registration for parametric global motion models," *Journal of Visual Communication and Image Representation*, vol. 23, no. 7, pp. 1060–1067, 2012.

[5] S. Villena, M. Vega, D. Babacan, R. Molina, and A. Katsaggelos, "Bayesian combination of sparse and non sparse priors in image super resolution," *Digital Signal Processing*, vol. 23, no. 2, pp. 530–541, 2013.

[6] Chia-Kai Liang, Li-Wen Chang, and Homer H Chen, "Analysis and compensation of rolling shutter effect," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1323–1330, 2008.

[7] Simon Baker, Eric Bennett, Sing Bing Kang, and Richard Szeliski, "Removing rolling shutter wobble," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2392–2399.

[8] Matthias Grundmann, Vivek Kwatra, Daniel Castro, and Irfan Essa, "Calibration-free rolling shutter removal," in *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2012, pp. 1–8.

[9] Erik Ringaby and Per-Erik Forssén, "Efficient video rectification and stabilisation for cell-phones," *International Journal of Computer Vision*, vol. 96, no. 3, pp. 335–352, 2012.

[10] VijayRengarajanAngarai Pichaikuppan, RajagopalanAmbasamudram Narayanan, and Aravind Rangarajan, "Change detection in the presence of motion blur and rolling shutter effect," in *Computer Vision - ECCV 2014*, vol. 8695 of *LNCS*, pp. 123–137. Springer, 2014.

[11] Shuochen Su and Wolfgang Heidrich, "Rolling shutter motion deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1529–1537.

[12] Johan Hedborg, Per-Erik Forssén, Michael Felsberg, and Erik Ringaby, "Rolling shutter bundle adjustment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 1434–1441.

[13] Sunghoon Im, Hyowon Ha, Gyeongmin Choe, Hae-Gon Jeon, Kyungdon Joo, and In So Kweon, "High quality structure from small motion for rolling shutter cameras," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 837–845.

[14] Gustav Hanning, Nicklas Forslöw, Per-Erik Forssén, Erik Ringaby, David Törnqvist, and Jonas Callmer, "Stabilizing cell phone video using inertial measurement sensors," in *International Conference on Computer Vision Workshops*. IEEE, 2011, pp. 1–8.

[15] Chao Jia and Brian L Evans, "Probabilistic 3-d motion estimation for rolling shutter video rectification from visual and inertial measurements.," in *International Workshop on Multimedia Signal Processing*, 2012, pp. 203–208.

[16] Ondfej Sindelar, Filip Sroubek, and Peyman Milanfar, "A smartphone application for removing handshake blur and compensating rolling shutter," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2160–2162.

[17] Abhijith Punnappurath, Vijay Rengarajan, and A.N. Rajagopalan, "Rolling shutter super-resolution," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 558–566.

[18] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[19] Ken Shoemake, "Animating rotation with quaternion curves," in *ACM SIGGRAPH Computer Graphics*. ACM, 1985, vol. 19, pp. 245–254.

[20] F. Sroubek, G. Cristobal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2322–2332, Sept 2007.