Image Stitching and Rectification for Hand-Held Cameras

Bingbing Zhuang[0000−0002−2317−3882] and Quoc-Huy Tran[0000−0003−1396−6544]

NEC Labs America

Abstract. In this paper, we derive a new differential homography that can account for the scanline-varying camera poses in Rolling Shutter (RS) cameras, and demonstrate its application to carry out RS-aware image stitching and rectification at one stroke. Despite the high complexity of RS geometry, we focus in this paper on a special yet common input — two consecutive frames from a video stream, wherein the inter-frame motion is restricted from being arbitrarily large. This allows us to adopt simpler differential motion model, leading to a straightforward and practical minimal solver. To deal with non-planar scene and camera parallax in stitching, we further propose an RS-aware spatially-varying homogarphy field in the principle of As-Projective-As-Possible (APAP). We show superior performance over state-of-the-art methods both in RS image stitching and rectification, especially for images captured by hand-held shaking cameras.

Keywords: Rolling shutter, Image rectification, Image stitching, Differential homography, Homography field, Hand-held cameras

1 Introduction

Rolling Shutter (RS) cameras adopt CMOS sensors due to their low cost and simplicity in manufacturing. This stands in contrast to Global Shutter (GS) CCD cameras that require specialized and highly dedicated fabrication. Such discrepancy endows RS cameras great advantage for ubiquitous employment in consumer products, e.g., smartphone cameras [44] or dashboard cameras [12]. However, the expediency in fabrication also causes a serious defect in image capture — instead of capturing different scanlines all at once as in GS cameras, RS cameras expose each scanline one by one sequentially from top to bottom. While static RS camera capturing a static scene is fine, the RS effect comes to haunt us as soon as images are taken during motion, i.e., images could be severely distorted due to scanline-varying camera poses (see Fig. 1).

RS distortion has been rearing its ugly head in various computer vision tasks. There is constant pressure to either remove the RS distortion in the front-end image capture [25, 48, 50, 62], or design task-dependent RS-aware algorithms in the back end [54, 15, 10, 2, 42, 51, 46]. While various algorithms have been developed for each of them in isolation, algorithms achieving both in a holistic way

Fig. 1. Example results of RS-aware image stitching and rectification.

are few [52, 67, 24, 63]. In this paper, we make contributions towards further advancement in this line. Specifically, we propose a novel differential homography and demonstrate its application to carry out RS image stitching and rectification at one stroke.

RS effect complicates the two-view geometry significantly compared to its GS counterpart, primarily because 12 additional unknown parameters are required to model the intra-frame velocity of the two cameras. Thus, despite the recent effort of Lao et al. [24] in solving a generic RS homography for discrete motion, the complexity of RS geometry significantly increases the number of required correspondences (36 points for full model and 13.5 points after a series of approximations). Inspired by prior work [67] that demonstrates dramatic simplification in differential RS relative pose estimation compared to its discrete counterpart [10], we focus in this paper on the special yet common case where the inputs are two consecutive frames from a video. In this case, the inter-frame motion is restricted from being arbitrarily large, allowing us to adopt the simpler differential homography model [39]. Furthermore, the intra-frame motion could be directly parameterized by the inter-frame motion via interpolation, thereby reducing the total number of unknown parameters to solve. In particular, we derive an RS-aware differential homography under constant acceleration motion assumption, together with a straightforward solver requiring only 5 pairs of correspondences, and demonstrate its application to simultaneous RS image stitching and rectification. Since a single homography warping is only exact under pure rotational camera motion or for 3D planar scene, it often causes misalignment when such condition is not strictly met in practice. To address such model inadequacy, we extend the single RS homography model to a spatially-varying RS homography field following the As-Projective-As-Possible (APAP) principle [64], thereby lending itself to handling complex scenes. We demonstrate example results in Fig. 1, where multiple images are stitched and rectified by concatenating pairwise warping from our method.

We would also like to emphasize our advantage over the differential Structure-from-Motion (SfM)-based rectification method [67]. Note that [67] computes the rectification for each pixel separately via pixel-wise depth estimation from optical

flow and camera pose. As such, potential gross errors in optical flow estimates could lead to severe artifacts in the texture-less or non-overlapping regions. In contrast, the more parsimonious homography model offers a natural defense against wrong correspondences. Despite its lack of full 3D reconstruction, we observe good empirical performance in terms of visual appearance.

In summary, our contributions include:

- We derive a novel differential homography model together with a minimal solver to account for the scanline-varying camera poses of RS cameras.
- We propose an RS-aware spatially-varying homography field for improving RS image stitching.
- Our proposed framework outperforms state-of-the-art methods both in RS image rectification and stitching.

2 Related Work

RS Geometry. Since the pioneering work of Meingast et al. [41], considerable efforts have been invested in studying the geometry of RS cameras. These include relative pose estimation [10, 67, 47], absolute pose estimation [40, 2, 56, 26, 3, 23], bundle adjustment [15, 22], SfM/Reconstruction [54, 20, 55, 58, 59], degeneracies [4, 21, 69], discrete homography [24], and others [5, 45]. In this work, we introduce RS-aware differential homography, which is of only slighly higher complexity than its GS counterpart.

RS Image Rectification. Removing RS artifacts using a *single* input image is inherently an ill-posed problem. Works in this line [50, 48, 25] often assume simplified camera motions and scene structures, and require line/curve detection in the image, if available at all. Recent methods [49, 69] have started exploring deep learning for this task. However, their generalization ability to different scenes remains an open problem. In contrast, *multi-view* approaches, be it geometric-based or learning-based [35], are more geometrically grounded. In particular, Ringaby and Forssen [52] estimate and smooth a sequence of camera rotations for eliminating RS distortions, while Grundmann et al. [11] and Vasu et al. [62] use a mixture of homographies to model and remove RS effects. Such methods often rely on nontrivial iterative optimization leveraging a large set of correspondences. Recently, Zhuang et al. [67] present the first attempt to derive minimal solver for RS rectification. It takes a minimal set of points as input and lends itself well to RANSAC, leading to a more principled way for robust estimation. In the same spirit, we derive RS-aware differential homography and show important advantages. Note that our minimal solver is orthogonal to the optimization-based methods, e.g. [52, 62], and can serve as their initialization. Very recently, Albl et al. [1] present an interesting way for RS undistortion from two cameras, yet require specific camera mounting.

GS Image Stitching. Image stitching [60] has achieved significant progress over the past few decades. Theoretically, a single homography is sufficient to align two input images of a common scene if the images are captured with no parallax or the scene is planar [13]. In practice, this condition is often violated, causing

misalignments or ghosting artifacts in the stitched images. To overcome this issue, several approaches have been proposed such as spatially-varying warps [34, 33, 64, 27, 29], shape-preserving warps [7, 8, 30], and seam-driven methods [66, 31, 17, 18]. All of the above approaches assume a GS camera model and hence they cannot handle RS images, i.e., the stitched images may contain RS distortion-induced misalignment. While Lao et al. [24] demonstrate the possibility of stitching in spite of RS distortion, we present a more concise and straightforward method that works robustly with hand-held cameras.

3 Homography Preliminary

GS Discrete Homography. Let us assume that two calibrated cameras are observing a 3D plane parameterized as $(\boldsymbol{n}, d)$, with $\boldsymbol{n}$ denoting the plane normal and d the camera-to-plane distance. Denoting the relative camera rotation and translation as $\boldsymbol{R} \in SO(3)$ and $\boldsymbol{t} \in \mathbb{R}^3$, a pair of 2D correspondences $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (in normalized plane) can be related by $\hat{\boldsymbol{x}}_2 \propto \boldsymbol{H}\hat{\boldsymbol{x}}_1$, where $\boldsymbol{H} = \boldsymbol{R} + \boldsymbol{t}\boldsymbol{n}^\top/d$ is defined as the *discrete* homography [13] and $\hat{\boldsymbol{x}} = [\boldsymbol{x}^\top, 1]^\top$. $\propto$ indicates equality up to a scale. Note that $\boldsymbol{H}$ in the above format subsumes the pure rotation-induced homography as a special case by letting $d \to \infty$. Each pair of correspondence $\{\boldsymbol{x}_1^i, \boldsymbol{x}_2^i\}$ gives two constraints $\boldsymbol{a}_i\boldsymbol{h} = \boldsymbol{0}$, where $\boldsymbol{h} \in \mathbb{R}^9$ is the vectorized form of $\boldsymbol{H}$ and the coefficients $\boldsymbol{a}_i \in \mathbb{R}^{2\times 9}$ can be computed from $\{\boldsymbol{x}_1^i, \boldsymbol{x}_2^i\}$. In *GS discrete 4-point solver*, with the minimal of 4 points, one can solve $\boldsymbol{h}$ via:

$$\boldsymbol{A}\boldsymbol{h} = \boldsymbol{0}, \quad s.t.\ \|\boldsymbol{h}\| = 1, \tag{1}$$

which has a closed-form solution by Singular Value Decomposition (SVD). $\boldsymbol{A}$ is obtained by stacking all $\boldsymbol{a}_i$.

GS Spatially-Varying Discrete Homography Field. In image stitching application, it is often safe to make zero-parallax assumption as long as the (non-planar) scene is far enough. However, it is also not uncommon that such assumption is violated to the extent that warping with just one global homography causes unpleasant misalignments. To address this issue, APAP [64] proposes to compute a spatially-varying homograpy field for each pixel $\boldsymbol{x}$:

$$\boldsymbol{h}^*(\boldsymbol{x}) = \arg\min_{\boldsymbol{h}} \sum_{i\in\mathcal{I}} \|w_i(\boldsymbol{x})\boldsymbol{a}_i\boldsymbol{h}\|^2, \quad s.t.\ \|\boldsymbol{h}\| = 1, \tag{2}$$

where $w_i(\boldsymbol{x}) = \max(\exp(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{\sigma^2}), \tau)$ is a weight. σ and τ are the pre-defined scale and regularization parameters respectively. $\mathcal{I}$ indicates the inlier set returned from GS discrete 4-point solver with RANSAC (motivated by [61]). The optimization has a closed-form solution by SVD. On the one hand, Eq. 2 encourages the warping to be globally As-Projective-As-Possible (APAP) by making use of all the inlier correspondences, while, on the other hand, it allows local deformations guided by nearby correspondences to compensate for model deficiency. Despite being a simple tweak, it yet leads to considerable improvement in image stitching.

GS Differential Homography. Suppose the camera is undergoing an instantaneous motion [19], consisting of rotational and translational velocity $(\boldsymbol{\omega}, \boldsymbol{v})$. It would induce a motion flow $\boldsymbol{u} \in \mathbb{R}^2$ in each image point $\boldsymbol{x}$. Denoting $\tilde{\boldsymbol{u}} = [\boldsymbol{u}^\top, 0]^\top$, we have[1]

$$\tilde{\boldsymbol{u}} = (\boldsymbol{I} - \hat{\boldsymbol{x}}\boldsymbol{e}_3^\top)\boldsymbol{H}\hat{\boldsymbol{x}}, \tag{3}$$

where $\boldsymbol{H} = -(\lfloor\boldsymbol{\omega}\rfloor_\times + \boldsymbol{v}\boldsymbol{n}^\top/d)$ is defined as the *differential* homography [39]. $\boldsymbol{I}$ represents identity matrix and $\boldsymbol{e}_3 = [0, 0, 1]^\top$. $\lfloor . \rfloor_\times$ returns the corresponding skew-symmetric matrix from the vector. Each flow estimate $\{\boldsymbol{u}_i, \boldsymbol{x}_i\}$ gives two effective constraints out of the three equations included in Eq. 3, denoted as $\boldsymbol{b}_i\boldsymbol{h} = \boldsymbol{u}_i$, where $\boldsymbol{b}_i \in \mathbb{R}^{2\times 9}$ can be computed from $\boldsymbol{x}_i$. In *GS differential 4-point solver*, with a minimal of 4 flow estimates, $\boldsymbol{H}$ can be computed by solving:

$$\boldsymbol{B}\boldsymbol{h} = \boldsymbol{U}, \tag{4}$$

which admits closed-form solution by pseudo inverse. $\boldsymbol{B}$ and $\boldsymbol{U}$ are obtained by stacking all $\boldsymbol{b}_i$ and $\boldsymbol{u}_i$, respectively. Note that, we can only recover $\boldsymbol{H}_L = \boldsymbol{H} + \varepsilon\boldsymbol{I}$ with an unknown scale ε, because $\boldsymbol{B}$ has a one-dimensional null space. One can easily see this by replacing $\boldsymbol{H}$ in Eq. 3 with $\varepsilon\boldsymbol{I}$ and observing that the right hand side vanishes, regardless of the value of $\boldsymbol{x}$. ε can be determined subsequently by utilizing the special structure of calibrated $\boldsymbol{H}$. However, this is not relevant in our paper since we focus on image stitching on general uncalibrated images.

4 Methods

4.1 RS Motion Parameterization

Under the discrete motion model, in addition to the 6-Degree of Freedom (DoF) inter-frame relative motion $(\boldsymbol{R}, \boldsymbol{t})$, 12 additional unknown parameters $(\boldsymbol{\omega}_1, \boldsymbol{v}_1)$ and $(\boldsymbol{\omega}_2, \boldsymbol{v}_2)$ are needed to model the intra-frame camera velocity, as illustrated in Fig. 2(a). This quickly increases the minimal number of points and the algorithm complexity to compute an RS-aware homography. Instead, we aim to solve for the case of continuous motion, i.e., a relatively small motion between two consecutive frames. In this case, we only need to parameterize the relative motion $(\boldsymbol{\omega}, \boldsymbol{v})$ between the two first scanlines (one can choose other reference scanlines without loss of generality) of the image pair, and the poses corresponding to all the other scanlines can be obtained by interpolation, as illustrated in Fig. 2(b). In particular, it is shown in [67] that a *quadratic* interpolation can be derived under constant *acceleration* motion. Formally, the absolute camera rotation and translation $(\boldsymbol{r}_1^{y_1}, \boldsymbol{p}_1^{y_1})$ (resp. $(\boldsymbol{r}_2^{y_2}, \boldsymbol{p}_2^{y_2})$) of scanline y_1 (resp. y_2) in frame 1 (resp. 2) can be written as:

$$\boldsymbol{r}_1^{y_1} = \beta_1(k, y_1)\boldsymbol{\omega}, \quad \boldsymbol{p}_1^{y_1} = \beta_1(k, y_1)\boldsymbol{v}, \tag{5}$$

$$\boldsymbol{r}_2^{y_2} = \beta_2(k, y_2)\boldsymbol{\omega}, \quad \boldsymbol{p}_2^{y_2} = \beta_2(k, y_2)\boldsymbol{v}, \tag{6}$$

[1] See our supplementary material for derivations.

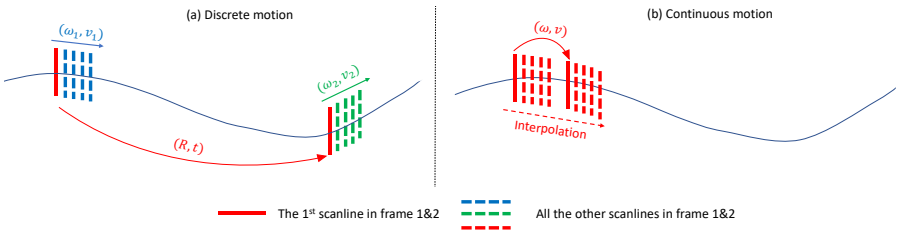

Fig. 2. Illustration of discrete/continuous camera motion and their motion parameters.

where

$$\beta_1(k, y_1) = (\frac{\gamma y_1}{h} + \frac{1}{2}k(\frac{\gamma y_1}{h})^2)(\frac{2}{2+k}), \tag{7}$$

$$\beta_2(k, y_2) = (1 + \frac{\gamma y_2}{h} + \frac{1}{2}k(1 + \frac{\gamma y_2}{h})^2)(\frac{2}{2+k}). \tag{8}$$

Here, k is an extra unknown motion parameter describing the acceleration, which is assumed to be in the same direction as velocity. γ denotes the the readout time ratio [67], i.e. the ratio between the time for scanline readout and the total time between two frames (including inter-frame delay). h denotes the total number of scanlines in a image. Note that the absolute poses $(\boldsymbol{r}_1^{y_1}, \boldsymbol{p}_1^{y_1})$ and $(\boldsymbol{r}_2^{y_2}, \boldsymbol{p}_2^{y_2})$ are all defined w.r.t the first scanline of frame 1. It follows that the relative pose between scanlines y_1 and y_2 reads:

$$\boldsymbol{\omega}_{y_1 y_2} = \boldsymbol{r}_2^{y_2} - \boldsymbol{r}_1^{y_1} = (\beta_2(k, y_2) - \beta_1(k, y_1))\boldsymbol{\omega}, \tag{9}$$

$$\boldsymbol{v}_{y_1 y_2} = \boldsymbol{p}_2^{y_2} - \boldsymbol{p}_1^{y_1} = (\beta_2(k, y_2) - \beta_1(k, y_1))\boldsymbol{v}. \tag{10}$$

We refer the readers to [67] for the detailed derivation of the above equations.

4.2 RS-Aware Differential Homography

We are now in a position to derive the RS-aware differential homography. First, it is easy to verify that Eq. 3 also applies uncalibrated cameras, under which case $\boldsymbol{H} = -\boldsymbol{K}(\lfloor\boldsymbol{\omega}\rfloor_\times + \boldsymbol{v}\boldsymbol{n}^\top/d)\boldsymbol{K}^{-1}$, with $\boldsymbol{u}$ and $\boldsymbol{x}$ being raw measurements in pixels. $\boldsymbol{K}$ denotes the unknown camera intrinsic matrix. Given a pair of correspondence by $\{\boldsymbol{u}, \boldsymbol{x}\}$, we can plug $(\boldsymbol{\omega}_{y_1 y_2}, \boldsymbol{v}_{y_1 y_2})$ into Eq. 3, yielding

$$\tilde{\boldsymbol{u}} = (\beta_2(k, y_2) - \beta_1(k, y_1))(\boldsymbol{I} - \hat{\boldsymbol{x}}\boldsymbol{e}_3^\top)\boldsymbol{H}\hat{\boldsymbol{x}} = \beta(k, y_1, y_2)(\boldsymbol{I} - \hat{\boldsymbol{x}}\boldsymbol{e}_3^\top)\boldsymbol{H}\hat{\boldsymbol{x}}. \tag{11}$$

Here, we can define $\boldsymbol{H}_{RS} = \beta(k, y_1, y_2)\boldsymbol{H}$ as the RS-aware differential homography, which is now scanline dependent.

5-Point Solver. In addition to $\boldsymbol{H}$, we now have one more unknown parameter k to solve. Below, we show that 5 pairs of correspondences are enough to solve for k and $\boldsymbol{H}$, using the so-called hidden variable technique [9]. To get started, let us first rewrite Eq. 11 as:

$$\beta(k, y_1, y_2)\boldsymbol{b}\boldsymbol{h} = \boldsymbol{u}. \tag{12}$$

Next, we move $\boldsymbol{u}$ to the left hand side and stack the constraints from 5 points, leading to:

$$\boldsymbol{C}\hat{\boldsymbol{h}} = \boldsymbol{0}, \tag{13}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \beta_1(k, y_1^1, y_2^1)\boldsymbol{b}_1, & -\boldsymbol{u}_1 \\ \beta_2(k, y_1^2, y_2^2)\boldsymbol{b}_2, & -\boldsymbol{u}_2 \\ \beta_3(k, y_1^3, y_2^3)\boldsymbol{b}_3, & -\boldsymbol{u}_3 \\ \beta_4(k, y_1^4, y_2^4)\boldsymbol{b}_4, & -\boldsymbol{u}_4 \\ \beta_5(k, y_1^5, y_2^5)\boldsymbol{b}_5, & -\boldsymbol{u}_5 \end{bmatrix}, \qquad \hat{\boldsymbol{h}} = [\boldsymbol{h}^T,\ 1]^T. \tag{14}$$

It is now clear that, for $\boldsymbol{h}$ to have a solution, $\boldsymbol{C}$ must be rank-deficient. Further observing that $\boldsymbol{C} \in \mathbb{R}^{10\times 10}$ is a square matrix, rank deficiency indicates vanishing determinate, i.e.,

$$det(\boldsymbol{C}) = \boldsymbol{0}. \tag{15}$$

This gives a univariable polynomial equation, whereby we can solve for k efficiently. $\boldsymbol{h}$ can subsequently be extracted from the null space of $\boldsymbol{C}$.

DoF Analysis. In fact, only 4.5 points are required in the minimal case, since we have one extra unknown k while each point gives two constraints. Utilizing 5 points nevertheless leads to a straightforward solution as shown. *Yet, does this lead to an over-constrained system*? No. Recall that we can only recover $\boldsymbol{H} + \varepsilon\boldsymbol{I}$ up to an arbitrary ε. Here, due to the one extra constraint, a specific value is chosen for ε since the last element of $\hat{\boldsymbol{h}}$ is set to 1. Note that a true ε, thus $\boldsymbol{H}$, is not required in our context since it does not affect the warping. This is in analogy to uncalibrated SfM [13] where a projective reconstruction up to an arbitrary projective transformation is not inferior to the Euclidean reconstruction in terms of reprojection error.

Plane Parameters. Strictly speaking, the plane parameters slightly vary as well due to the intra-frame motion. This is however not explicitly modeled in Eq. 11, due to two reasons. First, although the intra-frame motion is in a similar range as the inter-frame motion (Fig. 2(b)) and hence has a large impact in terms of motion, it induces merely a small perturbation to the absolute value of the scene parameters, which can be safely ignored (see supplementary for a more formal characterization). Second, we would like to keep the solver as simple as possible as along as good empirical results are obtained (see Sec. 5).

Motion Infidelity vs. Shutter Fidelity. Note that the differential motion model is always an approximation specially designed for small motion. This means that, unlike its discrete counterpart, its fidelity decreases with increasing motion. Yet, we are only interested in relatively large motion such that the RS distortion reaches the level of being visually unpleasant. Therefore, a natural and scientifically interesting question to ask is, whether the benefits from modeling RS distortion (Shutter Fidelity) are more than enough to compensate for the sacrifices due to the approximation in motion model (Motion Infidelity). Although a theoretical characterization on such comparison is out of the scope of this paper, via extensive experiments in Sec. 5, we fortunately observe that the differential RS model achieves overwhelming dominance in this competition.

Degeneracy. *Are there different pairs of k and $\boldsymbol{H}$ that lead to the same flow field $\boldsymbol{u}$?* Although such degeneracy does not affect stitching, it does make a difference to rectification (Sec. 4.4). We leave the detailed discussion to the supplementary, but would like the readers to be assured that such cases are very rare, in accordance with Horn [19] that motion flow is hardly ambiguous.
More Details. Firstly, note that although $\{\boldsymbol{u}, \boldsymbol{x}\}$ is typically collected from optical flow in classical works [38, 16] prior to the advent of keypoint descriptors (e.g., [37, 53]), we choose the latter for image stitching for higher efficiency. Secondly, if we fix $k = 0$, i.e., constant velocity model, $(\boldsymbol{\omega}, \boldsymbol{v})$ could be solved using a linear 4-point minimal solver similar to the GS case. However, we empirically find its performance to be inferior to the constant acceleration model in shaking cameras, and shall not be further discussed here.

4.3 RS-Aware Spatially-Varying Differential Homography Field

Can GS APAP [64] Handle RS Distortion by Itself? As aforementioned, the adaptive weight in APAP (Eq. 2) permits local deformations to account for the local discrepancy from the global model. However, we argue that APAP alone is still not capable of handling RS distortion. The root cause lies in the GS homography being used — although the warping of pixels near correspondences are less affected, due to the anchor points role of correspondences, the warping of other pixels still relies on the transformation propagated from the correspondences and thus the model being used does matter here.
RS-Aware APAP. Obtaining a set of inlier correspondences $\mathcal{I}$ from our RS differential 5-point solver with RANSAC, we formulate the spatially-varying RS-aware homography field as:

$$\boldsymbol{h}^*(\boldsymbol{x}) = \arg\min_{\boldsymbol{h}} \sum_{i \in \mathcal{I}} \|w_i(\boldsymbol{x})(\beta(k, y_1, y_2)\boldsymbol{b}_i\boldsymbol{h} - \boldsymbol{u}_i)\|^2, \tag{16}$$

where $w_i(\boldsymbol{x})$ is defined in Sec. 3. Since k is a pure motion parameter independent of the scene, we keep it fixed in this stage for simplicity. Normalization strategy [14] is applied to $(\boldsymbol{u}, \boldsymbol{x})$ for numerical stability. We highlight that the optimization has a simple closed-form solution, yet is geometrically meaningful in the sense that it minimizes the error between the estimated and the observed flow $\boldsymbol{u}$. This stands in contrast with the discrete homography for which minimizing reprojection error requires nonlinear iterative optimization. In addition, we also observe higher stability from the differential model in cases of keypoints concentrating in a small region (see supplementary for discussions).

4.4 RS Image Stitching and Rectification

Once we have the homography $\boldsymbol{H}$ (either a global one or a spatially-varying field) mapping from frame 1 to frame 2, we can warp between two images for stitching. Referring to Fig. 2(b) and Eq. 11, for each pixel $\boldsymbol{x}_1 = [x_1, y_1]^\top$ in frame 1, we find its mapping $\boldsymbol{x}_2 = [x_2, y_2]^\top$ in frame 2 by first solving for y_2 as:

$$y_2 = y_1 + \lfloor(\beta_2(k, y_2) - \beta_1(k, y_1))(\boldsymbol{I} - \hat{\boldsymbol{x}}_1\boldsymbol{e}_3^\top)\boldsymbol{H}\hat{\boldsymbol{x}}_1\rfloor_y, \tag{17}$$

which admits a closed-form solution. $\lfloor . \rfloor_y$ indicates taking the y coordinate. x_2 can be then obtained easily with known y_2. Similarly, $\boldsymbol{x}_1$ could also be projected to the GS canvas defined by the pose corresponding to the first scanline of frame 1, yielding its rectified point $\boldsymbol{x}_{g1}$. $\boldsymbol{x}_{g1}$ can be solved according to

$$\boldsymbol{x}_1 = \boldsymbol{x}_{g1} + \lfloor \beta_1(k, y_1)(\boldsymbol{I} - \hat{\boldsymbol{x}}_{g1}\boldsymbol{e}_3^\top)\boldsymbol{H}\hat{\boldsymbol{x}}_{g1} \rfloor_{xy}, \tag{18}$$

where $\lfloor . \rfloor_{xy}$ indicates taking x and y coordinate.

5 Experiments

5.1 Synthetic Data

Data Generation. First, we generate motion parameters $(\boldsymbol{\omega}, \boldsymbol{v})$ and k with desired constraints. For each scanline y_1 (resp. y_2) in frame 1 (resp. 2), we obtain its absolute pose as $(R(\beta_1(k, y_1)\boldsymbol{\omega}), \beta_1(k, y_1)\boldsymbol{v})$ (resp. $(R(\beta_2(k, y_2)\boldsymbol{\omega}), \beta_2(k, y_2)\boldsymbol{v})$). Here, $R(\boldsymbol{\theta}) = \exp(\lfloor \boldsymbol{\theta} \rfloor_\times)$ with exp: so(3) $\rightarrow$ SO(3). Due to the inherent depth-translation scale ambiguity, the magnitude of $\boldsymbol{v}$ is defined as the ratio between the translation magnitude and the average scene depth. The synthesized image plane is of size 720×1280 with a 60° horizontal Field Of View (FOV). Next, we randomly generate a 3D plane, on which we sample 100 3D points within FOV. Finally, we project each 3D point $\boldsymbol{X}$ to the RS image. Since we do not know which scanline observes $\boldsymbol{X}$, we first solve for y_1 from the quadratic equation:

$$y_1 = \lfloor \pi(\boldsymbol{R}(\beta_1(k, y_1)\boldsymbol{\omega})(\boldsymbol{X} - \beta_1(k, y_1)\boldsymbol{v})) \rfloor_y, \tag{19}$$

where $\pi([a, b, c]^\top) = [a/c, b/c]^\top$. x_1 can then be obtained easily with known y_1. Likewise, we obtain the projection in frame 2.

Comparison under Various Configurations. First, we study the performance under the noise-free case to understand the intrinsic and noise-independent behavior of different solvers, including discrete GS 4-point solver ('GS-disc'), differential GS 4-point solver ('GS-diff') and our RS 5-point solver ('RS-ConstAcc'). Specifically, we test the performance with varying RS readout time ratio γ, rotation magnitude $\|\boldsymbol{\omega}\|$, and translation magnitude $\|\boldsymbol{v}\|$. To get started, we first fix $(\|\boldsymbol{\omega}\|, \|\boldsymbol{v}\|)$ to $(3°, 0.03)$, and increase γ from 0 to 1, indicating zero to strongest RS effect. Then, we fix $\gamma = 1, \|\boldsymbol{v}\| = 0.03$ while increasing $\|\boldsymbol{\omega}\|$ from 0° to 9°. Finally, we fix $\gamma = 1, \|\boldsymbol{\omega}\| = 3°$ while increasing $\|\boldsymbol{v}\|$ from 0 to 0.1. We report averaged reprojection errors over all point pairs in Fig. 3(a)-(c). The curves are averaged over 100 configurations with random plane and directions of $\boldsymbol{\omega}$ and $\boldsymbol{v}$.

First, we observe that 'GS-diff' generally underperforms 'GS-disc' as expected due to its approximate nature (cf. 'Motion Infidelity' in Sec. 4.2). In (a), although 'RS-ConstAcc' performs slightly worse than 'GS-disc' under small RS effect ($\gamma <= 0.1$), it quickly surpasses 'GS-disc' significantly with increasing γ (cf. 'Shutter Fidelity' in Sec. 4.2). Moreover, this is constantly true in (b) and (c) with the gap becoming bigger with increasing motion magnitude. Such observations suggest that the gain due to handling RS effect overwhelms the degradation

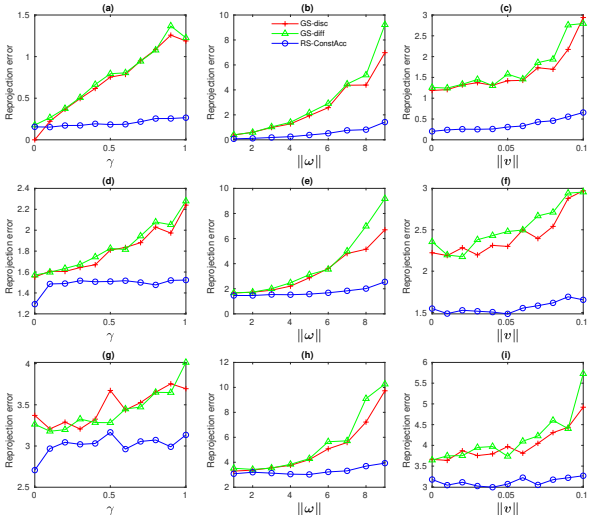

Fig. 3. Quantitative comparison under different configurations.

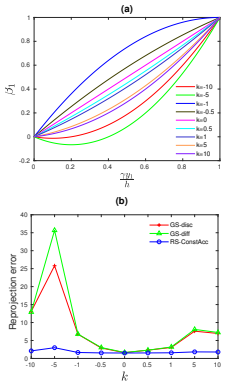

Fig. 4. Quantitative comparison under different values of k.

brought about by the less faithful differential motion model. Further, we conduct investigation with noisy data by adding Gaussian noise (with standard deviations $\sigma_g = 1$ and $\sigma_g = 2$ pixels) to the projected 2D points. The updated results in the above three settings are shown in Fig. 3(d)-(f) and Fig. 3(g)-(i) for $\sigma_g = 1$ and $\sigma_g = 2$ respectively. Again, we observe considerable superiority of the RS-aware model, demonstrating its robustness against noise. We also conduct evaluation under different values of k, with $(\|\boldsymbol{\omega}\|, \|\boldsymbol{v}\|) = (3°, 0.03)$, $\gamma = 1$, $\sigma_g = 1$. We plot $\beta_1(k, y_1)$ against $\frac{\gamma y_1}{h}$ with different values of k in Fig. 4(a) to have a better understanding of scanline pose interpolation. The reprojection error curves are plotted in Fig. 4(b). We observe that the performance of 'GS-disc' drops considerably with k deviating from 0, while 'RS-ConstAcc' maintains almost constant accuracy. Also notice the curves are not symmetric as $k>0$ indicates acceleration (increasing velocity) while $k<0$ indicates deceleration (decreasing velocity).

5.2 Real Data

We find that the RS videos used in prior works, e.g. [11, 52, 15], often contain small jitters without large viewpoint change across consecutive frames. To demonstrate the power of our method, we collect 5 videos (around 2k frames in total) with hand-held RS cameras while running, leading to large camera shaking and RS distortion. Following [35], we simply set $\gamma = 1$ to avoid its nontrival calibration [41] and find it works well for our camera.

Two-View Experiments. Below we discuss the two-view experiment results.

Qualitative Evaluation. We first present a few qualitative examples to intuitively demonstrate the performance gap, in terms of RS image rectification and stitching. For RS image rectification, we compare our method with the differential SfM

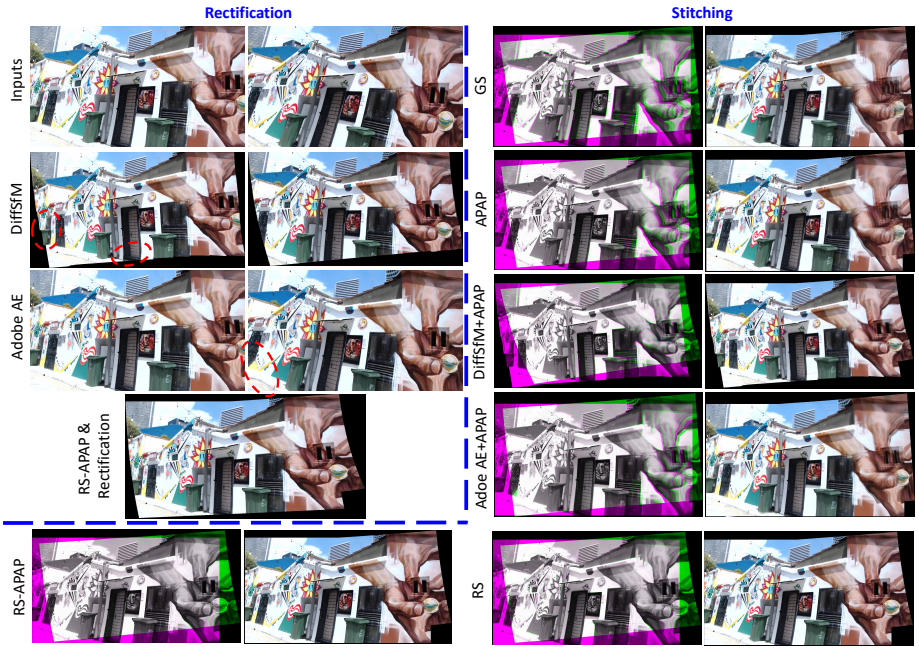

Fig. 5. Comparison of rectification/stitching on real RS images. Best viewed in screen.

based approach [67] ('DiffSfM') and the RS repair feature in Adobe After Effect ('Adobe AE'). For RS image stitching, we compare with the single GS discrete homography stitching ('GS') and its spatially-varying extension [64]('APAP'). In addition, we also evaluate the sequential approaches which feed 'DiffSfM' (resp. 'Adobe AE') into 'APAP', denoted as 'DiffSfM+APAP' (resp. 'Adobe AE+APAP'). We denote our single RS homography stitching without rectification as 'RS', our spatially-varying RS homography stitching without rectification as 'RS-APAP', and our spatially-varying RS homography stitching with rectification as 'RS-APAP & Rectification'.

In general, we observe that although 'DiffSfM' performs very well for pixels with accurate optical flow estimates, it may cause artifacts elsewhere. Similarly, we find 'Adobe AE' to be quite robust on videos with small jitters, but often introduces severe distortion with the presence of strong shaking. Due to space limit, we show two example results here and leave more to the supplementary.

In the example of Fig. 5, despite that 'DiffSfM' successfully rectifies the door and tube to be straight, the boundary parts (red circles) are highly skewed — these regions have no correspondences in frame 2 to compute flow. 'Adobe AE' manages to correct the images to some extent, yet bring evident distortion in the boundary too, as highlighted. 'RS-APAP & Rectification' nicely corrects the distortion with the two images readily stitched together. Regarding image stitching, we overlay two images after warping with the discrepancy visualized by

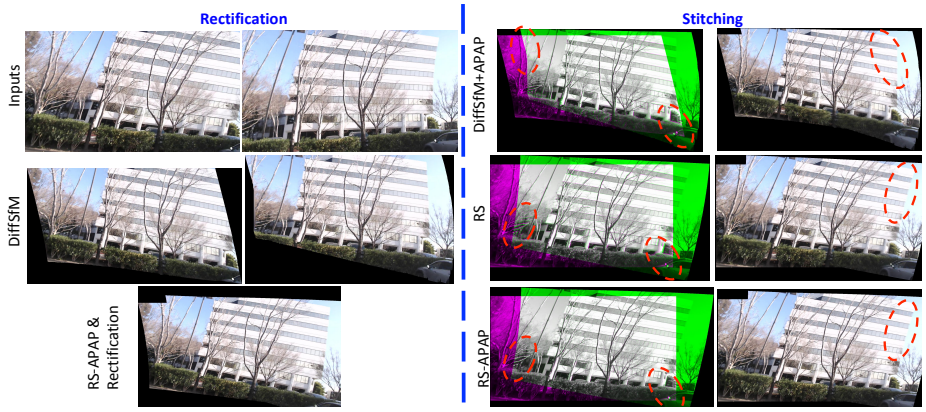

Fig. 6. Comparison of rectification/stitching on real RS images. Best viewed in screen.

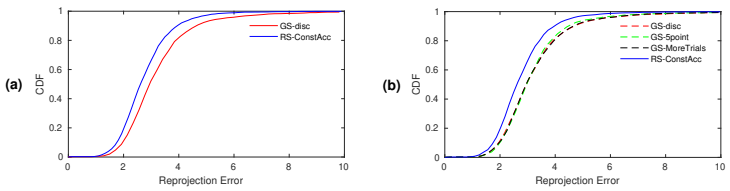

Fig. 7. Quantitative evaluation under standard setting & further study.

green/red colors, beside which we show the linearly blended images. As can be seen, 'GS' causes significant misalignments. 'APAP' reduces them to some extent but not completely. The artifacts due to 'DiffSfM' and 'Adobe AE' persist in the stitching stage. Even for those non-boundary pixels, there are still misalignments as the rectification is done per frame in isolation, independent of the subsequent stitching. In contrast, we observe that even one single RS homography ('RS') suffices to warp the images accurately here, yielding similar result as 'RS-APAP'.

We show one more example in Fig. 6 with partial results (the rest are in the supplementary). 'DiffSfM' removes most of the distortion to the extent that 'APAP' warps majority of the scene accurately ('DiffSfM+APAP'), yet, misalignments are still visible as highlighted, again, due to its sequential nature. We would like to highlight that APAP plays a role here to remove the misalignment left by the 'RS' and leads to the best stitching result.

Quantitative Evaluation. Here, we conduct quantitative evaluation to characterize the benefits brought about by our RS model. For every pair of consecutive frames, we run both 'GS-disc' and 'RS-ConstAcc', each with 1000 RANSAC trials. We compute for each pair the median reprojection error among all the correspondences, and plot its cumulative distribution function (CDF) across all the frame pairs, as shown in Fig. 7(a). Clearly, 'RS-ConstAcc' has higher-quality warping with reduced reprojection errors.

Table 1. RMSE evaluation for image stitching using different methods.

Method	GS	Mesh-based[36]	Mixture[11]	APAP[64]	RS-APAP
RMSE([0-255])	5.72	5.15	3.65	3.27	**3.05**

Although the above comparison demonstrates promising results in favor of the RS model, we would like to carry out further studies for more evidence, due to two reasons. First, note that the more complicated RS model has higher DoF and it might be the case that the smaller reprojection errors are simply due to over-fitting to the observed data, rather than due to truly higher fidelity of the underlying model. Second, different numbers (4 vs. 5) of correspondences are sampled in each RANSAC trial, leading to different amount of total samples used by the two algorithms. To address these concerns, we conduct two further investigations accordingly. First, for each image pair, we reserve 500 pairs of correspondences as test set and preclude them from being sampled during RANSAC. We then compare how well the estimated models perform on this set. Second, we test two different strategies to make the total number of samples equivalent — 'GS-MoreTrials': increases the number of RANSAC trials for 'GS-disc' to $1000\times5/4 = 1250$; 'GS-5point': samples non-minimal 5 points and get a solution in least squares sense in each trial. As shown in Fig. 7(b), although 'GS-5point' does improve the warping slightly, all the GS-based methods still lag behind the RS model, further validating the utility of our RS model.

Comparison with Homographies for Video Stabilization [36, 11]. Here, we compare with the mesh-based spatially-variant homographies [36] and the homography mixture [11] proposed for video stabilization. We would like to highlight that the fundamental limitation behind [36, 11] lies in that the individual homography is still GS-based, whereas ours explicitly models RS effect. We follow [28, 32] to evaluate image alignment by the RMSE of one minus normalized cross correlation (NCC) over a neighborhood of 3×3 window for the overlapping pixel $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, i.e. $RMSE = \sqrt{\frac{1}{N}\sum_{\pi}(1 - NCC(\boldsymbol{x}_i, \boldsymbol{x}_j))^2}$, with N being the total number of pixels in the overlapping region π. As shown in Table. 1, RS-APAP achieves lower averaged RMSE than [36, 11]. Surprisingly, [36] is not significantly better than GS, probably as its shape-preserving constraint becomes too strict for our strongly shaky videos. We also note that, in parallel with MDLT, our RS model could be integrated into [36, 11] as well; this is however left as future works.

Test on Data from [67]. We also compare with [24, 62] on the 6 image pairs used in [67], with 2 shown in Fig. 8 and 4 in the supplementary. We show the results from our single RS model without APAP for a fair comparison to [67, 24]. First, we observe that our result is not worse than the full 3D reconstuction method [67]. In addition, it can be seen that our method performs on par with [24, 62], while being far more concise and simpler.

Multiple-View Experiments. We demonstrate an extension to multiple images by concatenating the pairwise warping (note that the undertermined ε's do

Fig. 8. Qualitative comparison to DiffSfM [67], the method of Lao and Ait-Aider [24], and the method of Vasu et al. [62]. Stitched images with rectification are shown for [24] and ours.

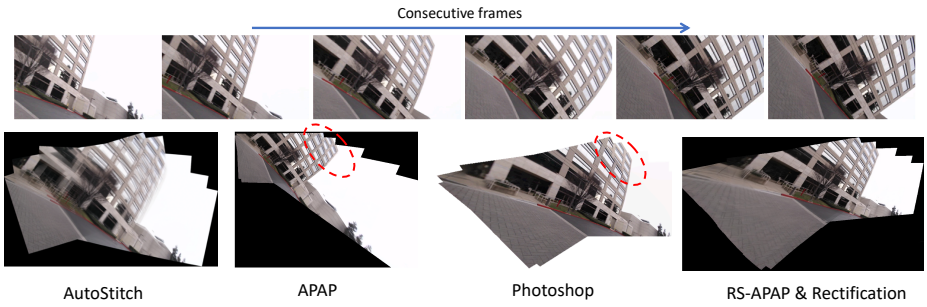

Fig. 9. Qualitative comparison on multiple image stitching.

not affect this step). We show an example in Fig. 9 and compare with the multi-image APAP [65], AutoStitch [6] and Photoshop. AutoStitch result exhibits severe ghosting effects. APAP repairs them but not completely. Photoshop applies advanced seam cutting for blending, yet can not obscure the misalignments. Despite its naive nature, our simple concatenation shows superior stitching results.

6 Conclusion

We propose a new RS-aware differential homography together with its spatially-varying extension to allow local deformation. At its core is a novel minimal solver strongly governed by the underlying RS geometry. We demonstrate its application to RS image stitching and rectification at one stroke, achieving good performance. We hope this work could shed light on handling RS effect in other vision tasks such as large-scale SfM/SLAM [43, 57, 68, 4, 70].

Acknowledgements. We would like to thank Buyu Liu, Gaurav Sharma, Samuel Schulter, and Manmohan Chandraker for proofreading and support of this work. We are also grateful to all the reviewers for their constructive suggestions.

References

1. Albl, C., Kukelova, Z., Larsson, V., Polic, M., Pajdla, T., Schindler, K.: From two rolling shutters to one global shutter. In: CVPR (2020)
2. Albl, C., Kukelova, Z., Pajdla, T.: R6p-rolling shutter absolute camera pose. In: CVPR (2015)
3. Albl, C., Kukelova, Z., Pajdla, T.: Rolling shutter absolute pose problem with known vertical direction. In: CVPR (2016)
4. Albl, C., Sugimoto, A., Pajdla, T.: Degeneracies in rolling shutter sfm. In: ECCV (2016)
5. Bapat, A., Price, T., Frahm, J.M.: Rolling shutter and radial distortion are features for high frame rate multi-camera tracking. In: CVPR (2018)
6. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision **74**(1), 59–73 (2007)
7. Chang, C.H., Sato, Y., Chuang, Y.Y.: Shape-preserving half-projective warps for image stitching. In: CVPR (2014)
8. Chen, Y.S., Chuang, Y.Y.: Natural image stitching with the global similarity prior. In: ECCV (2016)
9. Cox, D.A., Little, J., O'shea, D.: Using algebraic geometry, vol. 185. Springer Science & Business Media (2006)
10. Dai, Y., Li, H., Kneip, L.: Rolling shutter camera relative pose: Generalized epipolar geometry. In: CVPR (2016)
11. Grundmann, M., Kwatra, V., Castro, D., Essa, I.: Calibration-free rolling shutter removal. In: ICCP (2012)
12. Haresh, S., Kumar, S., Zia, M.Z., Tran, Q.H.: Towards anomaly detection in dashcam videos. In: IV (2020)
13. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press (2003)
14. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Transactions on pattern analysis and machine intelligence **19**(6), 580–593 (1997)
15. Hedborg, J., Forssén, P.E., Felsberg, M., Ringaby, E.: Rolling shutter bundle adjustment. In: CVPR (2012)
16. Heeger, D.J., Jepson, A.D.: Subspace methods for recovering rigid motion i: Algorithm and implementation. International Journal of Computer Vision **7**(2), 95–117 (1992)
17. Herrmann, C., Wang, C., Strong Bowen, R., Keyder, E., Krainin, M., Liu, C., Zabih, R.: Robust image stitching with multiple registrations. In: ECCV (2018)
18. Herrmann, C., Wang, C., Strong Bowen, R., Keyder, E., Zabih, R.: Object-centered image stitching. In: ECCV (2018)
19. Horn, B.K.: Motion fields are hardly ever ambiguous. International Journal of Computer Vision **1**(3), 259–274 (1988)
20. Im, S., Ha, H., Choe, G., Jeon, H.G., Joo, K., So Kweon, I.: High quality structure from small motion for rolling shutter cameras. In: ICCV (2015)
21. Ito, E., Okatani, T.: Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In: CVPR (2017)
22. Klingner, B., Martin, D., Roseborough, J.: Street view motion-from-structure-from-motion. In: ICCV (2013)
23. Kukelova, Z., Albl, C., Sugimoto, A., Pajdla, T.: Linear solution to the minimal absolute pose rolling shutter problem. In: ACCV (2018)

24. Lao, Y., Aider, O.A.: Rolling shutter homography and its applications. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
25. Lao, Y., Ait-Aider, O.: A robust method for strong rolling shutter effects correction using lines with automatic feature selection. In: CVPR (2018)
26. Lao, Y., Ait-Aider, O., Bartoli, A.: Rolling shutter pose and ego-motion estimation using shape-from-template. In: ECCV (2018)
27. Lee, K.Y., Sim, J.Y.: Warping residual based image stitching for large parallax. In: CVPR (2020)
28. Li, S., Yuan, L., Sun, J., Quan, L.: Dual-feature warping-based motion model estimation. In: ICCV (2015)
29. Liao, T., Li, N.: Single-perspective warps in natural image stitching. IEEE Transactions on Image Processing **29**, 724–735 (2019)
30. Lin, C.C., Pankanti, S.U., Natesan Ramamurthy, K., Aravkin, A.Y.: Adaptive as-natural-as-possible image stitching. In: CVPR (2015)
31. Lin, K., Jiang, N., Cheong, L.F., Do, M., Lu, J.: Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In: ECCV (2016)
32. Lin, K., Jiang, N., Liu, S., Cheong, L.F., Do, M., Lu, J.: Direct photometric alignment by mesh deformation. In: CVPR (2017)
33. Lin, W.Y., Liu, S., Matsushita, Y., Ng, T.T., Cheong, L.F.: Smoothly varying affine stitching. In: CVPR (2011)
34. Liu, F., Gleicher, M., Jin, H., Agarwala, A.: Content-preserving warps for 3d video stabilization. ACM Transactions on Graphics (TOG) **28**(3), 1–9 (2009)
35. Liu, P., Cui, Z., Larsson, V., Pollefeys, M.: Deep shutter unrolling network. In: CVPR (2020)
36. Liu, S., Yuan, L., Tan, P., Sun, J.: Bundled camera paths for video stabilization. ACM Transactions on Graphics (TOG) **32**(4), 1–10 (2013)
37. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
38. Ma, Y., Košecká, J., Sastry, S.: Linear differential algorithm for motion recovery: A geometric approach. International Journal of Computer Vision **36**(1), 71–89 (2000)
39. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An invitation to 3-d vision: from images to geometric models, vol. 26. Springer Science & Business Media (2012)
40. Magerand, L., Bartoli, A., Ait-Aider, O., Pizarro, D.: Global optimization of object pose and motion from a single rolling shutter image with automatic 2d-3d matching. In: ECCV (2012)
41. Meingast, M., Geyer, C., Sastry, S.: Geometric models of rolling-shutter cameras. In: Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (2005)
42. Mohan, M.M., Rajagopalan, A., Seetharaman, G.: Going unconstrained with rolling shutter deblurring. In: ICCV (2017)
43. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: Orb-slam: a versatile and accurate monocular slam system. IEEE transactions on robotics **31**(5), 1147–1163 (2015)
44. Muratov, O., Slynko, Y., Chernov, V., Lyubimtseva, M., Shamsuarov, A., Bucha, V.: 3dcapture: 3d reconstruction for a smartphone. In: CVPRW (2016)
45. Oth, L., Furgale, P., Kneip, L., Siegwart, R.: Rolling shutter camera calibration. In: CVPR (2013)
46. Punnappurath, A., Rengarajan, V., Rajagopalan, A.: Rolling shutter super-resolution. In: ICCV (2015)
47. Purkait, P., Zach, C.: Minimal solvers for monocular rolling shutter compensation under ackermann motion. In: WACV (2018)

48. Purkait, P., Zach, C., Leonardis, A.: Rolling shutter correction in manhattan world. In: ICCV (2017)
49. Rengarajan, V., Balaji, Y., Rajagopalan, A.: Unrolling the shutter: Cnn to correct motion distortions. In: CVPR (2017)
50. Rengarajan, V., Rajagopalan, A.N., Aravind, R.: From bows to arrows: Rolling shutter rectification of urban scenes. In: CVPR (2016)
51. Rengarajan, V., Rajagopalan, A.N., Aravind, R., Seetharaman, G.: Image registration and change detection under rolling shutter motion blur. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(10), 1959–1972 (2016)
52. Ringaby, E., Forssén, P.E.: Efficient video rectification and stabilisation for cell-phones. International Journal of Computer Vision **96**(3), 335–352 (2012)
53. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: ICCV (2011)
54. Saurer, O., Koser, K., Bouguet, J.Y., Pollefeys, M.: Rolling shutter stereo. In: ICCV (2013)
55. Saurer, O., Pollefeys, M., Hee Lee, G.: Sparse to dense 3d reconstruction from rolling shutter images. In: CVPR (2016)
56. Saurer, O., Pollefeys, M., Lee, G.H.: A minimal solution to the rolling shutter pose estimation problem. In: IROS (2015)
57. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: CVPR (2016)
58. Schubert, D., Demmel, N., Usenko, V., Stuckler, J., Cremers, D.: Direct sparse odometry with rolling shutter. In: ECCV (2018)
59. Schubert, D., Demmel, N., Usenko, V., Stuckler, J., Cremers, D.: Direct sparse odometry with rolling shutter. In: ECCV (2018)
60. Szeliski, R., et al.: Image alignment and stitching: A tutorial. Foundations and Trends® in Computer Graphics and Vision **2**(1), 1–104 (2007)
61. Tran, Q.H., Chin, T.J., Carneiro, G., Brown, M.S., Suter, D.: In defence of ransac for outlier rejection in deformable registration. In: ECCV (2012)
62. Vasu, S., Mohan, M.M., Rajagopalan, A.: Occlusion-aware rolling shutter rectification of 3d scenes. In: CVPR (2018)
63. Vasu, S., Rajagopalan, A.N., Seetharaman, G.: Camera shutter-independent registration and rectification. IEEE Transactions on Image Processing **27**(4), 1901–1913 (2017)
64. Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. In: CVPR (2013)
65. Zaragoza, J., Chin, T.J., Tran, Q.H., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(7), 1285–1298 (2014)
66. Zhang, F., Liu, F.: Parallax-tolerant image stitching. In: CVPR (2014)
67. Zhuang, B., Cheong, L.F., Hee Lee, G.: Rolling-shutter-aware differential sfm and image rectification. In: ICCV (2017)
68. Zhuang, B., Cheong, L.F., Hee Lee, G.: Baseline desensitizing in translation averaging. In: CVPR (2018)
69. Zhuang, B., Tran, Q.H., Ji, P., Cheong, L.F., Chandraker, M.: Learning structure-and-motion-aware rolling shutter correction. In: CVPR (2019)
70. Zhuang, B., Tran, Q.H., Lee, G.H., Cheong, L.F., Chandraker, M.: Degeneracy in self-calibration revisited and a deep learning solution for uncalibrated slam. In: IROS (2019)