This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Scanline Homographies for Rolling-Shutter Plane Absolute Pose

Fang Bai*†, Agniva Sengupta*, Adrien Bartoli*
ENCOV, TGI, Institut Pascal, UMR6602 CNRS, Université Clermont Auvergne, France
fang.bai@yahoo.com, i.agniva@gmail.com, adrien.bartoli@gmail.com

Abstract

Cameras on portable devices are manufactured with a rolling-shutter (RS) mechanism, where the image rows (aka. scanlines) are read out sequentially. The unknown camera motions during the imaging process cause the so-called RS effects which are solved by motion assumptions in the literature. In this work, we give a solution to the absolute pose problem free of motion assumptions. We categorically demonstrate that the only requirement is motion smoothness instead of stronger constraints on the camera motion. To this end, we propose a novel mathematical abstraction for RS cameras observing a planar scene, called the scanline-homography, a 3×2 matrix with 5 DOFs. We establish the relationship between a scanline-homography and the corresponding plane-homography, a 3×3 matrix with 6 DOFs assuming the camera is calibrated. We estimate the scanline-homographies of an RS frame using a smooth image warp powered by B-Splines, and recover the plane-homographies afterwards to obtain the scanline-poses based on motion smoothness. We back our claims with various experiments. Code and new datasets: https://bitbucket.org/clermontferrand/planarscanlinehomography/src/master/.

1. Introduction

Many portable devices use cameras based on CMOS sensors equipped with a rolling-shutter (RS) mechanism, owing to its low-power consumption, low-price and high frame rate. In RS cameras, each scanline is exposed and read out sequentially [14] during imaging. This is different from cameras based on CCD sensors which mostly use a global-shutter (GS) mechanism that reads out all pixel rows at the same time. The RS mechanism becomes troublesome when imaging highly dynamic environments or when the camera itself undergoes high-speed motions when reading out those scanlines. In general, the effects are two-folds:

1) An RS camera violates the conventional pinhole camera model used in most vision algorithms [17], as most of them are derived based on the GS and the pinhole camera model [16].

2) An RS camera can generate images with unintended artefacts, *e.g.,* distortion and blur.

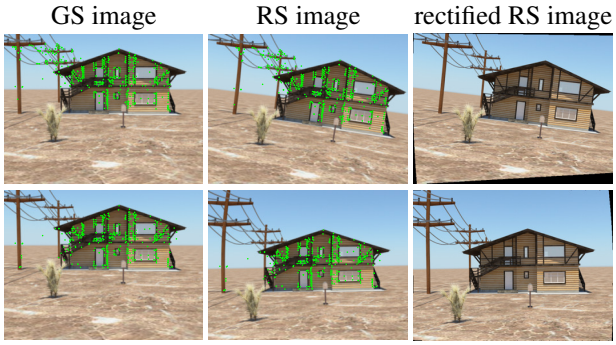

Figure 1. RS image rectification using a GS image as the template.

The first effect leads to research in RS geometry [9, 14, 26, 43] and the second effect to methods for RS image rectification [15, 27, 38, 40]; see section 2 for a review. All these works, with a few exceptions [28], are based on certain types of motion assumptions, *e.g.,* kinematic models [6, 9, 26, 43, 45], trajectory fitting with polynomials [35, 37, 40] and B-Splines [13, 21, 30, 33, 34], pose interpolation [12, 17, 38], or homography mixtures [15].

In this paper, we tackle the *RS absolute pose problem*, that is to estimate all the scanline poses from a single RS image given a template and *RS image rectification*. We specifically consider the case of planar objects, where the *template* is a 2D Euclidean object model [1, 2, 4, 6, 23, 31]. This problem admits a closed-form solution by using a constant kinematic motion model [1]. However, as we shall show in this paper, it is not trivial at all to solve the problem without motion assumptions. In this work, we only require the camera motion to be smooth; this stems from the RS image being a time-coherent set of scanlines rather than jagged fragments. Aside from this, we do not impose any assumption on the actual type of motion the camera undergoes, *i.e.*,

*Equal contribution, †Corresponding author

the motion can be simple kinematics, complicated dynamics, or even random movements as arbitrary as noise. In addition, we extend the template class to include the case of a GS image photoed by a camera whose principal axis is orthogonal to the planar scene (see Fig. 1) which is easy to obtain in practice.

Recently, there has been work [28] attempting to explain an RS image as a deformed GS image, based on isometric and conformal shape-from-template (SfT) methods [8]. The underlying idea is that as the RS goes along a certain direction, say the y-direction from top to bottom, the deformation along the x-direction preserves relative distances, resulting in a conformal deformation constraint along the x-direction. However, a general (isometric or conformal) SfT method constrains both x- and y-directions, causing the RS motion to be overconstrained.

To address the above issues, we propose a novel geometric concept for RS cameras, termed *scanline homography*. The set of scanline homographies of an RS image defines an image warp, with arbitrary flexibility along the y-direction (the RS direction) which admits all possible camera motions. The x-direction of the warp is constrained by the only admissible geometry, that the y-th scanline is imaged from a line in the template plane. The main contribution of this work is to establish the theory of scanline homographies with the following keypoints:

- We propose the scanline homography, a novel RS geometric concept given as a 3×2 matrix, which maps points on one RS scanline to a line in the template plane.
- We establish the fundamental homography equation between a scanline homography and its corresponding plane homography defined from the template to the RS image.
- We estimate scanline homographies parameterised by B-Splines along the RS direction to accommodate flexible camera motions by solving a convex minimisation in closed-form.
- We propose a method to solve the ambiguous fundamental homography equation to recover scanline poses, using an articulated plane homography parameterisation and a GS homography initialisation inspired from motion smoothness.
- We design various experiments to demonstrate the validity of the proposed theory, ranging from absolute pose estimation to image rectification.

2. Related work

RS geometry. Starting from the first attempt on RS geometric modeling by Geyer *et al.* [14], various works have been carried out to extend the classical GS multiple-view geometric techniques to the RS case, *e.g.,* the RS epipolar constraint [9], the RS homography constraint [26, 44], the RS optical flow equation [43], the RS bundle adjustment [17], the RS stereo-rig [5, 39] and the triangulation from an RS stereo-rig [3].

RS absolute pose. The RS absolute pose problem is to estimate scanline poses from a single RS image given a known template. When the template is a Euclidean point-cloud and the camera calibration is known, the problem is termed as the RS perspective-n-point (RS-PnP) problem [1, 2, 4, 6, 23, 31]. Typically, work in this regard assumes that the RS camera moves with constant linear and angular velocity during image acquisition. Ait-Aider *et al.* [1] proposed a projective RS-PnP formulation by minimizing the reprojection error using nonlinear least-squares. For planar objects, a closed-form solution was also given using homography constraints. An extension to line correspondences was given in [2]. Magerand *et al.* [31] proposed a globally optimal solution using a linearized Rodrigues formula for rotation parameterization. A set of polynomial equations are constructed and minimized by the Gloptipoly solver. There has been work to derive effective minimal RS-PnP solutions to be used in a RANSAC scheme. In specific, Albl *et al.* [6] proposed R6P, the first non-iterative minimal RS-PnP solver, using a double linearized RS model and the Gröbner basis solver to solve the polynomial equations. The efficiency of R6P is further improved in [23]. The Cayley parameterization for rotations was studied in [4]. Recently, the RS-PnP with an unknown focal length and radial distortion was considered in [24].

RS image rectification. There is a body of work dedicated specially to remove RS distortions, see [7,29] for early work on translation only camera motions. In the case of photoing using cellphones, the camera translation is normally negligible with respect to the scene depth, thus inspiring work based on rotation modeling only [15, 27, 38]. Ringaby *et al.* [38] parameterized the camera rotation using a smooth curve and SLERP (spherical linear interpolation), and estimated the parameters using nonlinear least-squares. Another representative work came from Grundmann *et al.* [15] who proposed a calibration-free method using a mixture of homographies. Su *et al.* [40] used image blur to recover both the latent GS image and camera motions which were parameterized with polynomials, in a unified formulation. Geometric information on the scene, *e.g.,* straightness of lines [37] and orthogonal vanishing directions [35] in a Manhattan world has been exploited. Lao *et al.* [27] improves over [35, 37] by first fitting parametric curves generated from a constant velocity model, then extracting the motion parameters from fitted curves in an RANSAC scheme. Occlusions were considered in [41] based on a multilayer 3D scene model. Data driven methods were proposed to

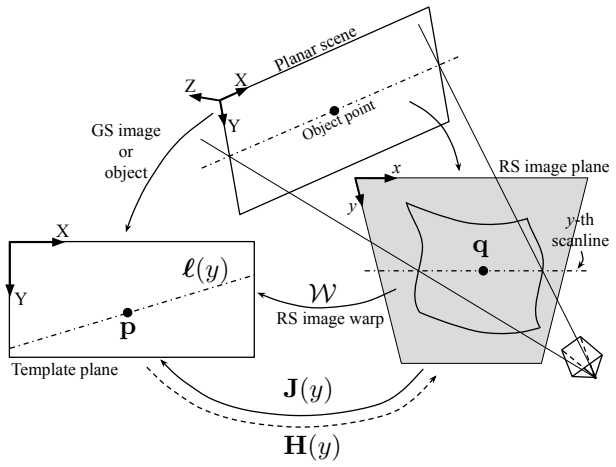

Figure 2. Scanline geometry for RS plane absolute pose.

learn motion kinematics [36], or depth as well to handle ambiguities [45] using a conventional neural network. These methods work on trained scenes but do not generalize to all scenarios. Recently, Albl *et al.* [5] showed how to obtain an undistorted GS image from an RS rig with different RS directions.

3. The Scanline Homography Model

3.1. General Points, Plane Homography

We consider the case where the object is planar, see figure 2. Our template can be both the planar Euclidean object or its image in a GS camera whose principal axis is orthogonal to the object plane. In both cases, a point in the template plane can be written as $[X;\, Y] \in \mathbb{R}^2$ and parameterized in the homogeneous form as $\mathbf{p} = [X;\, Y;\, 1] \in \mathbb{R}^3$. We denote the corresponding point in the RS image plane by $[x;\, y]$ and its homogeneous representation as $\mathbf{q} = [x;\, y;\, 1] \in \mathbb{R}^3$.

Without loss of generality, we stipulate the RS direction to be the y-axis. Thus, for each y-coordinate, the pixels along the x-coordinate are read out at the same time, which are subject to the same camera pose, termed the y-th scanline pose. For an RS image point $\mathbf{q}$ on the y-th RS scanline, $\mathbf{q}$ is related to its corresponding template point $\mathbf{p}$ by a plane homography [16]:

$$\mathbf{q} \propto \mathbf{H}(y)\,\mathbf{p}, \tag{1}$$

where $\propto$ means equality up to a non-zero scale. The plane-homography $\mathbf{H}(y)$ is defined as a mapping from the template frame to the RS image frame, thus can be decomposed to obtain the y-th scanline pose. Equation (1) is valid for all $\mathbf{q}$ on the y-th RS scanline, and is only valid for the y-th scanline.

3.2. Scanline Homography

Derivation. We examine the y-th scanline in the RS image, which is made of points imaged by the RS camera at the same time. The set of points in the template corresponding to this scanline is $\mathbf{p} \propto \mathbf{H}^{-1}(y)\,\mathbf{q}$. As each $\mathbf{q}$ has the same y-coordinate because it lies on the y-th scanline, we parameterize $\mathbf{q}$ with its x-coordinate only. For this purpose, we first denote the columns of $\mathbf{H}^{-1}(y)$ as $\mathbf{H}^{-1}(y) \propto [\mathbf{a}_1(y),\, \mathbf{a}_2(y),\, \mathbf{a}_3(y)]$, and write:

$$\mathbf{p} \propto \mathbf{H}^{-1}(y)\,\mathbf{q} \propto x\mathbf{a}_1(y) + y\mathbf{a}_2(y) + \mathbf{a}_3(y). \tag{2}$$

The above can be written compactly per scanline as:

$$\mathbf{p} \propto \mathbf{J}(y) \begin{bmatrix} x \\ 1 \end{bmatrix}, \tag{3}$$

where we introduce a novel matrix $\mathbf{J}(y)$, which we term the *scanline homography*:

$$\mathbf{J}(y) \stackrel{\text{def}}{=} [\mathbf{a}_1(y),\; y\mathbf{a}_2(y) + \mathbf{a}_3(y)] \stackrel{\text{def}}{=} [\mathbf{j}_1(y),\; \mathbf{j}_2(y)]. \tag{4}$$

The scanline homography matrix $\mathbf{J}(y)$ defines a mapping that brings an RS image point $\mathbf{q}$ on the y-th scanline to its corresponding point $\mathbf{p}$ in the template plane. Obviously, such a mapping is unique per scanline but varies across different scanlines. We formally denote this mapping, given by a matrix $\mathbf{J}(y) \in \mathbb{R}^{3\times 2}$ as the *scanline homography* of the y-th scanline. It is easy to see that a scanline homography is defined up to scale, thus has 5 DOFs.

Scanline geometry. The scanline homography $\mathbf{J}(y)$ maps the y-th scanline in the RS image to a line in the template plane given as $\boldsymbol{\ell}(y) = \mathbf{j}_1(y) \times \mathbf{j}_2(y)$, comprising points $x\mathbf{j}_1(y) + \mathbf{j}_2(y)$ parameterized by all x.

The 5 DOFs of a scanline homography. For a calibrated camera, the 5 DOFs of a scanline homography correspond to the 5 parameters of the scanline's partial pose. This can be understood geometrically. The line $\boldsymbol{\ell}(y)$ in the template plane is defined up to a 1-DOF rotational ambiguity, that is we can rotate $\boldsymbol{\ell}(y)$ arbitrarily along the line itself before applying $\mathbf{H}(y)$.

3.3. Fundamental Homography Equation

We introduce matrix $\mathbf{M}(y) = [\mathbf{j}_1(y),\, \mathbf{0},\, \mathbf{j}_2(y)]$, and rewrite equation (3) as:

$$\mathbf{J}(y) \begin{bmatrix} x \\ 1 \end{bmatrix} = \mathbf{M}(y)\mathbf{q} \propto \mathbf{p}. \tag{5}$$

Combining equations (1) and (5), we obtain:

$$\mathbf{H}(y)\mathbf{M}(y)\mathbf{q} \propto \mathbf{q}, \tag{6}$$

which is satisfied for each RS image point $\mathbf{q}$ on the y-th scanline. We rewrite this equation explicitly as:

$$(\mathbf{H}(y)\mathbf{M}(y) - s\mathbf{I})\mathbf{q} = \mathbf{0}, \tag{7}$$

where $s \neq 0$ is an unknown scalar. Expanding the above matrix form, we have:

$$[\mathbf{H}(y)\mathbf{j}_1(y) - s\mathbf{e}_1,\ -s\mathbf{e}_2,\ \mathbf{H}(y)\mathbf{j}_2(y) - s\mathbf{e}_3] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{0}, \quad (8)$$

which is satisfied for all x given y. This implies that $\mathbf{H}(y)\mathbf{j}_1(y) - s\mathbf{e}_1 = \mathbf{0}$. We thus obtain:

$$\begin{cases} \mathbf{H}(y)\mathbf{j}_1(y) - s\mathbf{e}_1 = \mathbf{0} \\ -ys\mathbf{e}_2 + \mathbf{H}(y)\mathbf{j}_2(y) - s\mathbf{e}_3 = \mathbf{0}. \end{cases} \quad (9)$$

By introducing a matrix:

$$\mathbf{N}(y) = [\mathbf{e}_1,\ y\mathbf{e}_2 + \mathbf{e}_3] = \begin{bmatrix} 1 & 0 \\ 0 & y \\ 0 & 1 \end{bmatrix}, \quad (10)$$

we arrive at the *fundamental homography equation*:

$$\mathbf{H}(y)\mathbf{J}(y) = s\mathbf{N}(y) \ \Leftrightarrow\ \mathbf{H}(y)\mathbf{J}(y) \propto \mathbf{N}(y). \quad (11)$$

This equation connects a scanline homography and the corresponding plane homography defined in the opposite direction. The fundamental homography equation (11) provides 5 constraints as it is defined up to scale.

3.4. Scanline Homography Estimation

A scanline homography $\mathbf{J}(y)$ is an element lying in the 5-dimensional projective space $\mathbb{P}^5$. If the underlying camera motion is smooth, $\mathbf{J}(y)$ changes smoothly across scanlines parameterized in y, forming a smooth curve in $\mathbb{P}^5$. This curve defines a smooth nonlinear image warp from the RS image plane to the template.

There exist many ways to parameterize a curve in the projective space. Here we present one as follow:

$$\mathbf{\Gamma}(y) : \mathbb{R} \to \mathbb{R}^5, \quad \mathbf{J}(y) = \begin{bmatrix} \gamma_1(y) & \gamma_4(y) \\ \gamma_2(y) & \gamma_5(y) \\ \gamma_3(y) & 1 \end{bmatrix}, \quad (12)$$

where $\mathbf{\Gamma}(y) = [\gamma_1(y), \gamma_2(y), \gamma_3(y), \gamma_4(y), \gamma_5(y)]^\top$ is a curve defined in $\mathbb{R}^5$, parameterized with respect to y. The bottom-right elements of $\mathbf{J}(y)$ is set to 1 to constrain the scale. We find it reasonable to do so because if that element vanished, an RS image point $\mathbf{q} = [0; y; 1]$ would be mapped to infinity in the template, which is an unfeasible event. The curve $\mathbf{\Gamma}(y)$ defined in $\mathbb{R}^5$ thus induces a curve in the projective space $\mathbb{P}^5$.

We parameterize $\mathbf{\Gamma}(y)$ using polynomials and B-Splines [10, 13, 21, 30, 33, 34], while other options are possible, *e.g.,* polynomials [35, 37, 40], interpolation [12, 17, 38] or mixture models [15] which were all heavily exploited. In any case, we assume each $\gamma_j(y)$ to be written as:

$$\gamma_j(y) = \boldsymbol{\phi}_j(y)^\top \mathbf{w}_j, \quad (13)$$

where $\boldsymbol{\phi}_j(y) \in \mathbb{R}^m$ denotes a collection of given basis functions, and $\mathbf{w}_j \in \mathbb{R}^m$ the weight parameters to be estimated.

Given a set of point correspondences $\{\mathbf{p} \leftrightarrow \mathbf{q}\}$, we estimate $\mathbf{\Gamma}(y)$ from the scanline homography definition given in equation (3). Based on parameterization (12), we formulate the following 'algebraic' cost [16]:

$$\min_{\mathbf{\Gamma}} \sum_{\mathbf{p}\leftrightarrow\mathbf{q}} \left\| \begin{bmatrix} x & & -xX & 1 & \\ & x & -xY & & 1 \end{bmatrix} \mathbf{\Gamma}(y) - \begin{bmatrix} X \\ Y \end{bmatrix} \right\|^2. \quad (14)$$

For $\mathbf{\Gamma}(y)$ in the form of equation (13), problem (14) is linear least-squares in $\mathbf{w}_j$ ($j = 1, \ldots, 5$), thus is solved in closed-form. In practice, we implemented 1) the linear least-squares solver for input correspondences free of mismatches and 2) an M-estimator for input correspondences potentially contaminated by mismatches.

3.5. Scanline Homographies as an Image Warp

The smooth scanline homographies $\mathbf{J}(\cdot)$ across all scanlines form a nonlinear image warp $\mathcal{W}$ from the RS image plane to the template plane:

$$\mathcal{W}([x;\ y;\ 1]) \overset{\text{def}}{=} \mathbf{J}(y) \begin{bmatrix} x \\ 1 \end{bmatrix} \ \text{for all}\ (x,\ y)\ \text{in an RS image.}$$

The warp $\mathcal{W}$ captures all possible image distortions caused by camera motions assuming $\mathbf{J}(\cdot)$ is flexible enough in the RS direction y. Note that so far, we have not imposed any assumption on the camera motion. This is fundamentally different from previous works [6, 9, 12, 13, 15, 17, 21, 26, 30, 33–35, 37, 38, 40, 43, 45], which make explicit assumptions on the possible motions at this stage.

The scanline geometry asserts that the y-th scanline corresponds to the line $\boldsymbol{\ell}(y) = \mathbf{j}_1(y) \times \mathbf{j}_2(y)$ in the template plane. This is the only geometry we can tell in an RS camera imaging process. In fact, any geometry across scanlines requires additional motion assumptions. This leads to an obvious flaw in previous isometry and conformity based RS image warps [28], where the warp is constrained in both x- and y-directions. Needless to say, the constraint along y-direction is invalid as it must obviously be motion independent.

At last, the RS image warp $\mathcal{W}$ induced by scanline homographies is constrained by the scanline geometry, that a scanline is photoed from a line. The isometric and conformal constraints are approximations to this scanline geometry, thus the warps based on isometry and conformity [8, 28] are as well approximations to $\mathcal{W}$ while they limit possible camera motions.

4. Scanline Pose Estimation

We denote the y-th scanline pose as $\mathbf{P}(y) = [\mathbf{R}(y),\ \mathbf{t}(y)]$, where $\mathbf{R}(y) = [\mathbf{r}_1(y),\ \mathbf{r}_2(y),\ \mathbf{r}_3(y)]$ is the

rotation and $\mathbf{t}(y)$ the translation. We assume the camera is calibrated, thus $\mathbf{H}(y)$ is the so-called Euclidean homography, with 6 DOFs [11, 32, 42]. We shall give two parameterizations of $\mathbf{H}(y)$ in section 4.1 from the scanline pose $\mathbf{P}(y)$.

From equation (11), $\mathbf{J}(y)$ provides only 5 constraints on $\mathbf{H}(y)$, thus we cannot solve for a unique $\mathbf{H}(y)$ from equation (11). We solve this ambiguity in Section 4.2 in an optimization formulation using an approximate GS initialization. The recovered $\mathbf{H}(y)$ satisfies equation (11) exactly, and lies in the vicinity of the initialization, which ensures smoothness.

4.1. Plane Homography Parameterization

The template is a Euclidean object. In this case, the scanline pose $\mathbf{P}(y)$ is defined in the object's XYZ-Euclidean coordinate frame. Without loss of generality, we assume the planar object is defined in the XY-plane (*i.e.,* $Z = 0$). The plane homography takes the form:

$$\mathbf{H}_{\text{Object}}(y) = [\mathbf{r}_1(y),\, \mathbf{r}_2(y),\, \mathbf{t}(y)]. \tag{15}$$

The template is a GS image. In this case, the scanline pose $\mathbf{P}(y)$ is defined with respect to the camera pose of the GS image. We consider the case where the principal axis of the GS camera is orthogonal to the planar Euclidean object. The plane homography is:

$$\mathbf{H}_{\text{Frame}}(y) = [\mathbf{r}_1(y),\, \mathbf{r}_2(y),\, \mathbf{r}_3(y) + \mathbf{t}(y)]. \tag{16}$$

A unified parameterization. By noticing that $\mathbf{r}_1(y), \mathbf{r}_2(y), \mathbf{r}_3(y)$ constitute a basis of the Euclidean coordinate frame, we can parameterize the translation $\mathbf{t}(y)$ in this frame and write $\mathbf{t}(y) = t_1\mathbf{r}_1(y) + t_2\mathbf{r}_2(y) + t_3\mathbf{r}_3(y)$. Following this idea, both the plane homographies $\mathbf{H}_{\text{Object}}(y)$ and $\mathbf{H}_{\text{Frame}}(y)$ can be written as:

$$\mathbf{H}(y) = [\mathbf{r}_1(y),\, \mathbf{r}_2(y),\, a\mathbf{r}_1(y) + b\mathbf{r}_2(y) + c\mathbf{r}_3(y)]. \tag{17}$$

The above $\mathbf{H}(y)$ has the following structure:

$$\mathbf{H}(y) = \mathbf{R}(y)\mathbf{S}(y), \text{ with } \mathbf{S}(y) = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & c \end{bmatrix}, \tag{18}$$

which is the product of a rotation matrix and an upper triangular matrix. Given $\mathbf{H}(y)$, the rotation component $\mathbf{R}(y)$ and the translation component $\mathbf{S}(y)$ can be obtained via the QR decomposition of $\mathbf{H}(y)$.

4.2. Scanline Pose Estimation

Using equation (18) into equation (11), we have:

$$\mathbf{R}(y)\mathbf{S}(y)\mathbf{J}(y) \propto \mathbf{N}(y), \tag{19}$$

which is the basic equation for scanline pose estimation. Unfortunately, equation (19) provides only 5 constraints while $\mathbf{R}(y)$ and $\mathbf{S}(y)$ contains 6 parameters in total. As a consequence, due to the missing 1 DOF, the recovery of $\mathbf{R}(y)$ and $\mathbf{S}(y)$ from equation (19) is not unique.

To proceed, we first rewrite the set of solutions of equation (19) as the optimal solutions to the following optimization problem:

$$\begin{cases} \min_{s(y),\mathbf{R}(y),\mathbf{S}(y)} & \|s(y)\mathbf{R}(y)\mathbf{S}(y)\mathbf{J}(y) - \mathbf{N}(y)\|_{\mathcal{F}}^2 \\ \text{s.t.} & \mathbf{R}(y)^\top\mathbf{R}(y) = 1, \quad \det(\mathbf{R}(y)) = 1, \end{cases} \tag{20}$$

where $s(y) \in \mathbb{R}$ is a scale factor to be determined. The optimal value of problem (20) is always zero, while the optimal solution is not unique. At first, it seems nothing has been achieved from equation (19) to problem (20). However, it gives us the possibility to apply various optimization techniques to further constrain the solution, *e.g.,* via initialization or regularization. We propose a method based on initialization, which solves equation (19) exactly in the vicinity of the initialization. We use the *approximate GS pose* as initialization, while other initializations (*e.g.,* those from constant velocity models) can be trivially used instead.

4.2.1 Approximate Global-Shutter Pose

If we ignore the RS effects, all the scanline poses are the same, corresponding to the *approximate GS pose*, where we denote the related plane homography by $\mathbf{H}_{\text{GS}} = \mathbf{R}_{\text{GS}}\mathbf{S}_{\text{GS}}$. In specific, $\mathbf{H}(y) = \mathbf{H}_{\text{GS}}$ for all y in a GS approximation. In this case, from equation (4), the first column of $\mathbf{J}(y)$ is independent of y, thus can be used to determine the scale.

We assume $\mathbf{J}(y)$ has been normalized by its first column so that $\|\mathbf{j}_1(y)\|_2 = 1$. Equation (11) in the GS approximation then becomes:

$$\mathbf{H}_{\text{GS}}\mathbf{J}(y) = \mathbf{N}(y), \quad \text{for each } y. \tag{21}$$

Given a set of scanlines $y_1, y_2, \ldots, y_n$, we can estimate $\mathbf{H}_{\text{GS}}$ in closed-form as:

$$\mathbf{H}_{\text{GS}} = \mathcal{J}\mathcal{N}^\dagger, \tag{22}$$

with $\mathcal{J} = [\mathbf{J}(y_1), \ldots, \mathbf{J}(y_n)]$, $\mathcal{N} = [\mathbf{N}(y_1), \ldots, \mathbf{N}(y_n)]$. We finally decompose $\mathbf{H}_{\text{GS}}$ to obtain $\mathbf{R}_{\text{GS}}$ and $\mathbf{S}_{\text{GS}}$.

4.2.2 Scanline Pose from Alternations

Starting from the approximate GS pose given as $\mathbf{R}_{\text{GS}}$, $\mathbf{S}_{\text{GS}}$, we alternate the estimation of $\mathbf{R}(y)$ and $\mathbf{S}(y)$ to find a solution to problem (20). This process involves the resolution of $\mathbf{R}(y)$ given $\mathbf{S}(y)$, and vice versa the resolution of $\mathbf{S}(y)$ given $\mathbf{R}(y)$. Fortunately both subproblems are solvable in closed-form.

Estimation of $s(y)$ and $\mathbf{R}(y)$ given $\mathbf{S}(y)$. This is the scaled special-orthogonal Procrustes problem [18–20] whose solution is:

$$\mathbf{R}(y) = \mathbf{V}\,\mathrm{diag}(1, 1, \det(\mathbf{V}\mathbf{U}^\top))\,\mathbf{U}^\top, \tag{23}$$

where $\mathbf{S}(y)\mathbf{J}(y)\mathbf{N}(y)^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is the Singular Value Decomposition (SVD) of $\mathbf{S}(y)\mathbf{J}(y)\mathbf{N}(y)^\top$. The optimal scale factor $s(y)$ is given as:

$$s(y) = \frac{\mathrm{tr}(\mathbf{R}(y)\mathbf{S}(y)\mathbf{J}(y)\mathbf{N}(y)^\top)}{\mathrm{tr}(\mathbf{S}(y)\mathbf{J}(y)\mathbf{J}(y)^\top\mathbf{S}(y)^\top)}. \tag{24}$$

Estimation of $\mathbf{S}(y)$ given $s(y)$ and $\mathbf{R}(y)$. This is linear least-squares whose solution is:

$$\mathbf{S}(y) = \mathbf{I} + \left(\frac{1}{s(y)}\mathbf{R}(y)^\top\mathbf{N}(y) - \mathbf{J}(y)\right)\mathbf{v}(y)^\dagger\,[0,\,0,\,1], \tag{25}$$

with $\mathbf{v}(y) = [0,\,0,\,1]\mathbf{J}(y)$, and $\mathbf{v}(y)^\dagger$ its Moore–Penrose pesudo-inverse.

Algorithm. Starting from $\mathbf{R}_{\mathrm{GS}}, \mathbf{S}_{\mathrm{GS}}$, we alternate between equations (23, 24) and (25) until convergence to obtain the y-th scanline pose. Importantly, we achieve the above results without any motion assumptions, but requiring the RS image to be similar to the GS image which is always true in practice by smoothness.

5. Image Rectification

From the previous presentation, we have given a solution to RS plane absolute pose without motion assumptions. The computed scanline poses can be used to rectify an RS image such that all its scanlines are arranged in the same camera pose. We choose the pose of an arbitrary scanline y_0 as the anchor, and denote its plane homography as $\mathbf{H}(y_0)$. Then for each RS image pixel coordinate $\mathbf{q}$ on the y-th scanline, we obtain its rectified image coordinate, denoted by $\mathbf{q}'$, as:

$$\mathbf{q}' \propto \mathbf{H}(y_0)\mathbf{p} \propto \mathbf{H}(y_0)\mathbf{H}(y)^{-1}\mathbf{q}. \tag{26}$$

We process all scanlines using equation (26) to obtain the rectified image. For images, as $\mathbf{q}'$ is fractional, we use equation (26) with interpolations based on a bilinear kernel.

6. Experimental Results

6.1. Pose Estimation

We use the synthetic dataset from [12], specifically the 12 RS images from the *house_trans_rot1_B40* sequence. The facade of the house is planar (see Fig. 1 for samples). We use the GS image frame09 as the template. This dataset is simulated by a constant velocity model across scanlines and provides ground-truth (GT) scanline poses.

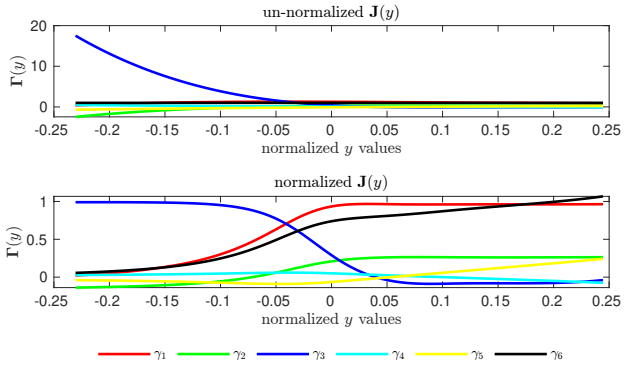

Figure 3. An example of estimated scanline-homography $\mathbf{J}(y)$. We use γ_6 to denote the bottom-right corner element in $\mathbf{J}(y)$, where $\gamma_6 = 1$ in equation (12) as shown on the left. For a calibrated camera, we know the first column of $\mathbf{J}(y)$ has unit norm, see Section 4.2.1. We thus normalize $\mathbf{J}(y)$ by the norm of its first column and show such normalized $\mathbf{J}(y)$ on the right.

The parameterization of scanline homography. For calibrated cameras, by equations (17) and (4), we know that the first column of $\mathbf{J}(y)$ admits unit norm. We thus first show in Fig. 3 that how this normalization constraint affects the complexity of the estimated scanline homography. This result shows how a simple $\mathbf{J}(y)$ parameterization can imply more complex normalized $\mathbf{J}(y)$. Therefore we suggest using simple parameterizations $\mathbf{J}(y)$ over advanced ones.

Scanline pose estimation and RS image rectification. We report the pose estimation error in Fig. 4 over two typical $\mathbf{J}(y)$ prameterizations, *i.e.,* polynomials (denoted by p.....) and B-splines (denoted by p.....c.....). The sequences after p indicate the degree of polynomials and c the number of control points in B-splines, respectively for each curve γ_j. The result shows that simple parameterizations *e.g.*, p11111, p11122, and p11111-c33333 are almost sufficient in most cases, while other complex parameterizations are prune to overfitting especially when the number of correspondences are low. We provide image rectification examples in Fig. 1, using the polynomial parameterization p11122. Although the scene is not strictly planar, we see that this does not substantially violate the planar model, as shown in the image rectification results.

6.2. Image Rectification

We provide additional examples on RS image rectification in Fig. 5, using the scanline homography parameterizations p11122 and p22222-c44444. Since existing researches are not dedicated to handle a GS image template explicitly, we use the RS plane relative pose solution [26] based on a constant velocity model as the benchmark algorithm where the template is considered as an RS image in [26].

We see that a specific absolute pose problem which handles a GS image template is useful, in particular when the

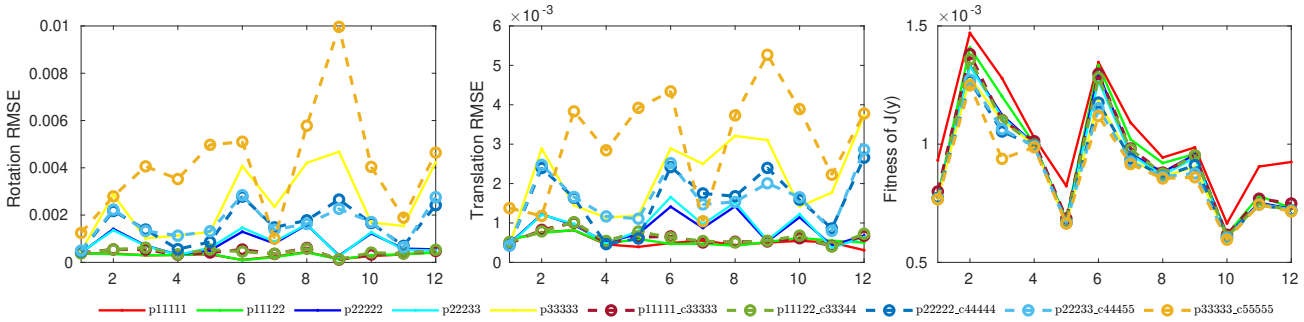

Figure 4. Scanline pose estimation error of the sequence *house_trans_rot1_B40* [12] which has 12 frames plotted along the x-axis. We use relative pose error (RPE) [25] to avoid gauge transformations. The sequence after 'p' indicates polynomial-degrees for each curve in $\mathbf{\Gamma}(y)$. The sequence after 'c' indicates the number of control points in the B-spline parameterization. The fitness of the scanline homography $\mathbf{J}(y)$ estimation is computed as the root mean squared error (RMSE) of the optimal cost of equation (14).

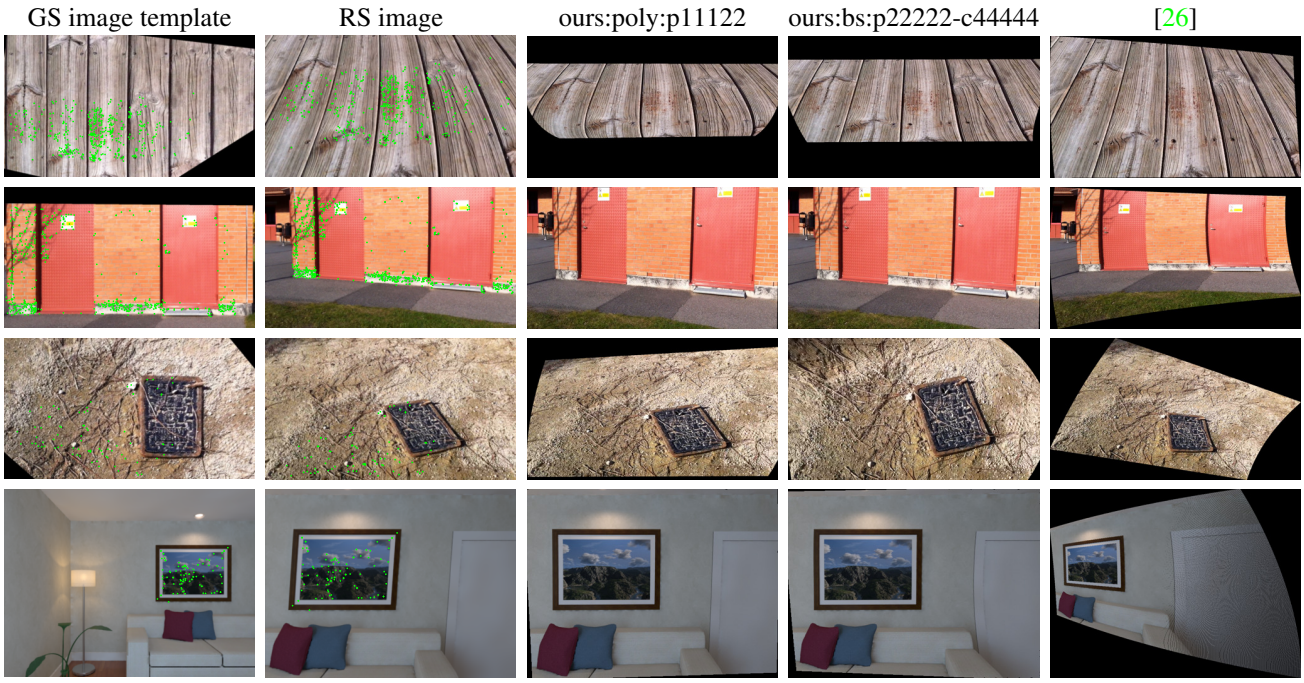

Figure 5. Rolling-shutter image rectification from a front-view global-shutter image template. We use handcrafted GS images for the first three rows (images from [17]) and a real GS image for the last row (images from [22]). We compare our RS plane absolute pose solution with the RS plane relative pose solution [26] based on a constant velocity model. The correspondences are plotted as green dots.

correspondences are sparse, *e.g.,* in the last two rows. With that being said, while a simple polynomial p11122 seems to be sufficient in most cases, an advanced B-spline parameterization can handle larger distortions, *e.g.,* the wooden floor in the first row rectified by p22222-c44444.

6.3. Handling Arbitrary Rolling-Shutter Motions

To test the limit of the proposed method, we use synthetic data generated with *arbitrary RS camera motions*.

Dataset generation. We simulate three challenging RS images using arbitrary motions in Fig. 6. Three different textures are mapped to a densely sampled planar surface, resulting in a trivial bijective map from the surface to the texture. Arbitrary smooth motions, generated in $\mathbb{R}^3$ using Bézier curves, are used to transform this template plane. A sequence of images depicting such motions are combined by each scanline to arrive at the final synthetic RS image.

Pose estimation. We report the scanline pose estimation error in Fig. 7, using the relative pose error as in Fig. 4. Interestingly, we find that the pose estimation results are strongly accurate for motion along X and Y axis, while some loss of accuracy remains for the motion along Z-axis (see the

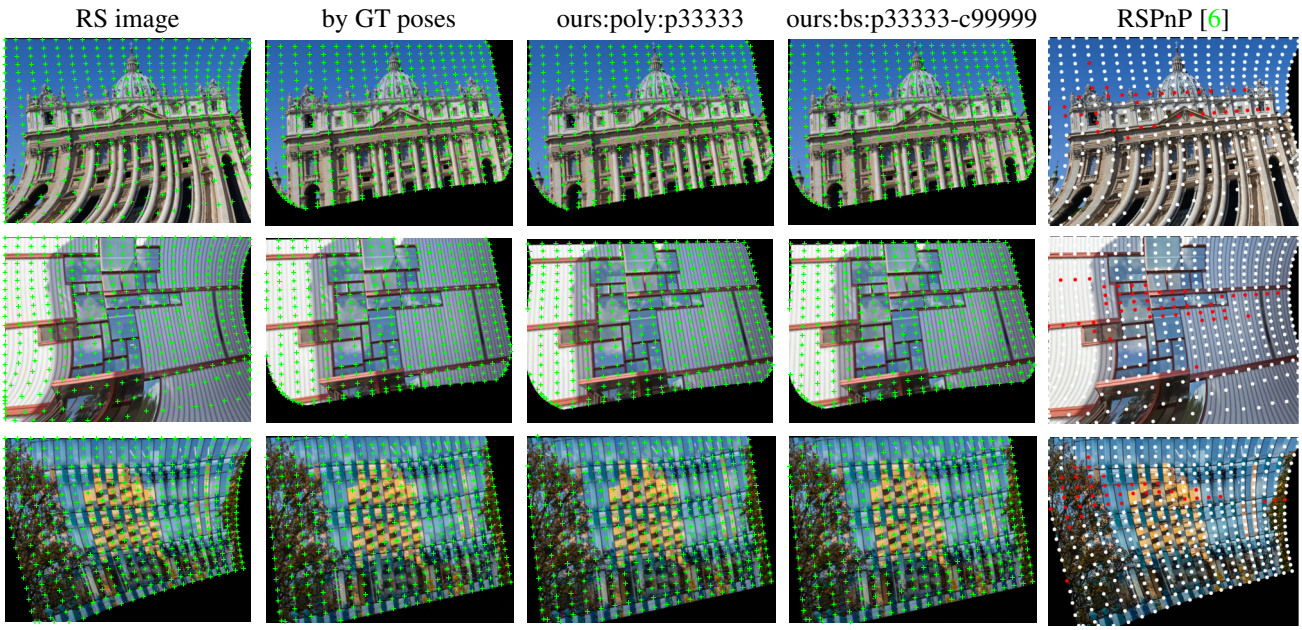

Figure 6. Rolling-shutter image rectification using the 3D object template. We simulate arbitrary camera motions and use a set of evenly distributed fiducial landmarks. The rectification result obtained using GT poses are given in the second column. We report two results of our method, *i.e.,* poly:p33333 and bs:p33333-c99999. We compare with the standard RS-PnP method [6] based on a linearized motion model. The RS-PnP method [6] is used within a RANSAC scheme, and the inliers admitted by the RS motion model are plotted in red.

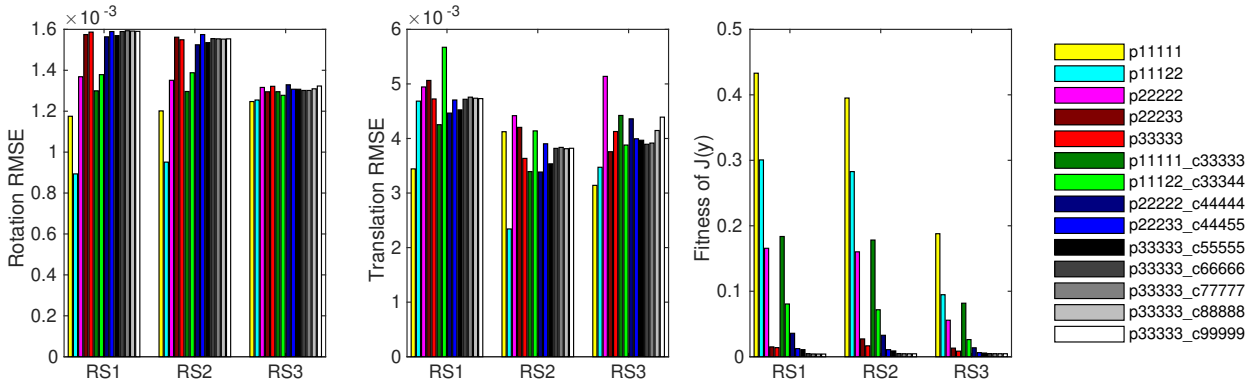

Figure 7. Scanline pose estimation error across all parameterizations using the same metrics as explained in Fig. 4.

supplementary material). However, this does not strongly impact image rectification accuracy.

Image rectification. We report image rectification results in Fig. 6. Despite the complex motions and the ridiculous RS distortions as a consequence, our method can give almost exactly the same rectification result compared with the GT rectification.

These above results confirm that the proposed method is capable to handle arbitrary motions, assuming the underlying distortion is captured by the correspondences. We release our code and the data for future comparisons. More results are provided in the supplementary material.

7. Conclusion

We have proposed the scanline homography, a novel mathematical abstraction for planar RS geometry. The scanline homography maps a scanline in the RS image to a line in the template plane. Such geometric constraints induce an image warp, which can be designed to be flexible enough along the RS direction to capture any camera motion. We establish key equations to recover the underlying plane homographies and thus the scanline poses. Experiments on absolute pose estimation and image rectification have validated our claims. Future work will extend our method to non-planar objects and relative pose estimation.

References

[1] Omar Ait-Aider, Nicolas Andreff, Jean Marc Lavest, and Philippe Martinet. Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In *European Conference on Computer Vision*, pages 56–68. Springer, 2006. 1, 2

[2] Omar Ait-Aider, Adrien Bartoli, and Nicolas Andreff. Kinematics from lines in a single rolling shutter image. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007. 1, 2

[3] Omar Ait-Aider and François Berry. Structure and kinematics triangulation with a rolling shutter stereo rig. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1835–1840. IEEE, 2009. 2

[4] Cenek Albl, Zuzana Kukelova, Viktor Larsson, and Tomas Pajdla. Rolling shutter camera absolute pose. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1439–1452, 2019. 1, 2

[5] Cenek Albl, Zuzana Kukelova, Viktor Larsson, Michal Polic, Tomas Pajdla, and Konrad Schindler. From two rolling shutters to one global shutter. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2505–2513, 2020. 2, 3

[6] Cenek Albl, Zuzana Kukelova, and Tomas Pajdla. R6P-rolling shutter absolute camera pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2292–2300, 2015. 1, 2, 4, 8

[7] Simon Baker, Eric Bennett, Sing Bing Kang, and Richard Szeliski. Removing rolling shutter wobble. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2392–2399. IEEE, 2010. 2

[8] Adrien Bartoli, Yan Gérard, Francois Chadebecq, Toby Collins, and Daniel Pizarro. Shape-from-template. *IEEE transactions on pattern analysis and machine intelligence*, 37(10):2099–2118, 2015. 2, 4

[9] Yuchao Dai, Hongdong Li, and Laurent Kneip. Rolling shutter camera relative pose: Generalized epipolar geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4132–4140, 2016. 1, 2, 4

[10] Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer-verlag New York, 1978. 4

[11] Olivier D Faugeras and Francis Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(03):485–508, 1988. 5

[12] Per-Erik Forssén and Erik Ringaby. Rectifying rolling shutter video from hand-held devices. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 507–514. IEEE, 2010. 1, 4, 6, 7

[13] Paul Furgale, Timothy D Barfoot, and Gabe Sibley. Continuous-time batch estimation using temporal basis functions. In *2012 IEEE International Conference on Robotics and Automation*, pages 2088–2095. IEEE, 2012. 1, 4

[14] Christopher Geyer, Marci Meingast, and Shankar Sastry. Geometric models of rolling-shutter cameras. *6th Workshop on Omnidirectional Vision*, 1:4, 2005. 1, 2

[15] Matthias Grundmann, Vivek Kwatra, Daniel Castro, and Irfan Essa. Calibration-free rolling shutter removal. In *2012 IEEE international conference on computational photography (ICCP)*, pages 1–8. IEEE, 2012. 1, 2, 4

[16] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 1, 3, 4

[17] Johan Hedborg, Per-Erik Forssén, Michael Felsberg, and Erik Ringaby. Rolling shutter bundle adjustment. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1434–1441. IEEE, 2012. 1, 2, 4, 7

[18] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *Josa a*, 4(4):629–642, 1987. 6

[19] Berthold KP Horn, Hugh M Hilden, and Shahriar Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, 5(7):1127–1135, 1988. 6

[20] Ken-ichi Kanatani. Analysis of 3-d rotation fitting. *IEEE Transactions on pattern analysis and machine intelligence*, 16(5):543–549, 1994. 6

[21] Christian Kerl, Jorg Stuckler, and Daniel Cremers. Dense continuous-time tracking and mapping with rolling shutter rgb-d cameras. In *Proceedings of the IEEE international conference on computer vision*, pages 2264–2272, 2015. 1, 4

[22] Jae-Hak Kim, Cesar Cadena, and Ian Reid. Direct semi-dense slam for rolling shutter cameras. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1308–1315. IEEE, 2016. 7

[23] Zuzana Kukelova, Cenek Albl, Akihiro Sugimoto, and Tomas Pajdla. Linear solution to the minimal absolute pose rolling shutter problem. In *Asian Conference on Computer Vision*, pages 265–280. Springer, 2018. 1, 2

[24] Zuzana Kukelova, Cenek Albl, Akihiro Sugimoto, Konrad Schindler, and Tomas Pajdla. Minimal rolling shutter absolute pose with unknown focal length and radial distortion. In *European Conference on Computer Vision*, pages 698–714. Springer, 2020. 2

[25] Rainer Kümmerle, Bastian Steder, Christian Dornhege, Michael Ruhnke, Giorgio Grisetti, Cyrill Stachniss, and Alexander Kleiner. On measuring the accuracy of slam algorithms. *Autonomous Robots*, 27(4):387–407, 2009. 7

[26] Yizhen Lao and Omar Ait Aider. Rolling shutter homography and its applications. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 1, 2, 4, 6, 7

[27] Yizhen Lao and Omar Ait-Aider. A robust method for strong rolling shutter effects correction using lines with automatic feature selection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4795–4803, 2018. 1, 2

[28] Yizhen Lao, Omar Ait-Aider, and Adrien Bartoli. Solving rolling shutter 3d vision problems using analogies with non-rigidity. *International Journal of Computer Vision*, 129(1):100–122, 2021. 1, 2, 4

[29] Chia-Kai Liang, Li-Wen Chang, and Homer H Chen. Analysis and compensation of rolling shutter effect. *IEEE Transactions on Image Processing*, 17(8):1323–1330, 2008. 2

[30] Steven Lovegrove, Alonso Patron-Perez, and Gabe Sibley. Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras. In *BMVC*, volume 2, page 8, 2013. 1, 4

[31] Ludovic Magerand, Adrien Bartoli, Omar Ait-Aider, and Daniel Pizarro. Global optimization of object pose and motion from a single rolling shutter image with automatic 2D-3D matching. In *European Conference on Computer Vision*, pages 456–469. Springer, 2012. 1, 2

[32] Ezio Malis and Manuel Vargas. *Deeper understanding of the homography decomposition for vision-based control*. PhD thesis, INRIA, 2007. 5

[33] Luc Oth, Paul Furgale, Laurent Kneip, and Roland Siegwart. Rolling shutter camera calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1360–1367, 2013. 1, 4

[34] Alonso Patron-Perez, Steven Lovegrove, and Gabe Sibley. A spline-based trajectory representation for sensor fusion and rolling shutter cameras. *International Journal of Computer Vision*, 113(3):208–219, 2015. 1, 4

[35] Pulak Purkait, Christopher Zach, and Ales Leonardis. Rolling shutter correction in Manhattan world. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 882–890, 2017. 1, 2, 4

[36] Vijay Rengarajan, Yogesh Balaji, and AN Rajagopalan. Unrolling the shutter: CNN to correct motion distortions. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 2291–2299, 2017. 3

[37] Vijay Rengarajan, Ambasamudram N Rajagopalan, and Rangarajan Aravind. From bows to arrows: Rolling shutter rectification of urban scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2773–2781, 2016. 1, 2, 4

[38] Erik Ringaby and Per-Erik Forssén. Efficient video rectification and stabilisation for cell-phones. *International journal of computer vision*, 96(3):335–352, 2012. 1, 2, 4

[39] Olivier Saurer, Kevin Koser, Jean-Yves Bouguet, and Marc Pollefeys. Rolling shutter stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 465–472, 2013. 2

[40] Shuochen Su and Wolfgang Heidrich. Rolling shutter motion deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1529–1537, 2015. 1, 2, 4

[41] Subeesh Vasu, AN Rajagopalan, et al. Occlusion-aware rolling shutter rectification of 3D scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 636–645, 2018. 2

[42] Zhongfei Zhang and Allen R Hanson. 3D reconstruction based on homography mapping. *Proc. ARPA96*, pages 1007–1012, 1996. 5

[43] Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Rolling-shutter-aware differential SfM and image rectification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 948–956, 2017. 1, 2, 4

[44] Bingbing Zhuang and Quoc-Huy Tran. Image stitching and rectification for hand-held cameras. In *European Conference on Computer Vision*, pages 243–260. Springer, 2020. 2

[45] Bingbing Zhuang, Quoc-Huy Tran, Pan Ji, Loong-Fah Cheong, and Manmohan Chandraker. Learning structure-and-motion-aware rolling shutter correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4551–4560, 2019. 1, 3, 4