

Rolling Shutter Homography and its Applications

Yizhen Lao and Omar Ait-Aider

Abstract—In this article we study the adaptation of the concept of homography to Rolling Shutter (RS) images. This extension has never been clearly addressed despite the many roles played by the homography matrix in multi-view geometry. We first show that a direct point-to-point relationship on a RS pair can be expressed as a set of 3 to 8 atomic 3×3 matrices depending on the kinematic model used for the instantaneous-motion during image acquisition. We call this group of matrices the RS Homography. We then propose linear solvers for the computation of these matrices using point correspondences. Finally, we derive linear and closed form solutions for two famous problems in computer vision in the case of RS images: image stitching and plane-based relative pose computation. Extensive experiments with both synthetic and real data from public benchmarks show that the proposed methods outperform state-of-art techniques.

Index Terms—Rolling shutter, Homography, Relative Pose estimation, Image stitching.

1 INTRODUCTION

A large number of consumer CMOS cameras are equipped with RS sensors because of their low-cost, low-energy-consumption, high frame rate and less-background-noise [1]. In this acquisition mode, image pixels are exposed row by row commonly from the top to the bottom. Thus, when the camera is moving during the acquisition, the exposure delay between the lines leads to visual distortions called RS effects, such as on Fig. 2(a).

Ignoring the effects of RS in computer vision applications results in performance degradation or even failure [2], [3]. Over the last decade, several works have revisited 3D computer vision by taking RS into account such as RS effects removal [4], [5], [6], [7], absolute pose estimation [2], [8], [9], [10], epipolar geometry [3], [11] and Structure from motion (SfM) [12], [13], [14], [15], [16], [17].

1.1 Related Works and Motivation

Estimating the camera motion by using points is one of the most studied minimal problems in computer vision. For example, with Global Shutter (GS), at least 3 point matches are needed to estimate the absolute pose [18], while at least 5 are needed to recover the relative pose between two calibrated GS views [19]. Given the higher complexity of RS projection model [20], more points are commonly needed. Methods for structure and motion estimation with RS images can be grouped into two categories: optical flow and epipolar geometry.

Optical flow methods: In [16], 8pt and 9pt linear solvers were developed to recover the relative pose of a RS camera that undergoes constant velocity or constant acceleration motion. Unfortunately, as shown in Fig. 1, consistency between the camera poses and their motion only holds with high-frame rates and smooth movements. In addition to the resulting high computation load, unordered image sets can not be processed. Besides, inter-frame delay has to be exactly pre-calibrated.

• Y. Lao and O. Ait-Aider are with Institut Pascal, Université Clermont Auvergne / CNRS, Clermont-Ferrand, France.
E-mail: lyz91822@gmail.com

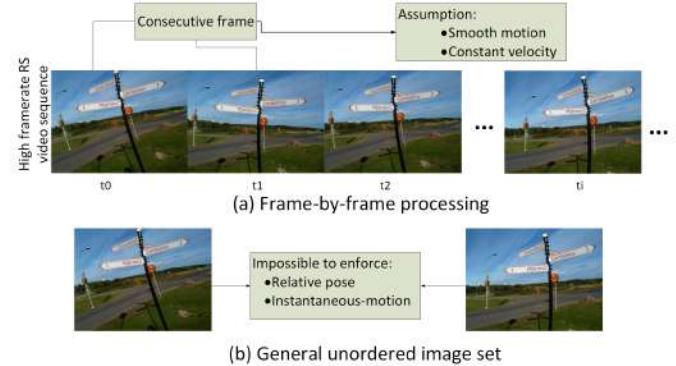


Fig. 1: (a) Video-based (frame-by-frame processing) methods assume smooth [4], [21] or even constant velocity [16] between each two consecutive frames. However, for a general unordered set of images (b), it is hard or impossible to enforce the relative poses and their instantaneous-motions basing on these assumptions.

Epipolar geometry: In multi-view reconstruction, many common configurations become critical with RS cameras and lead to reconstruction ambiguities. Authors in [15] provide mathematical analysis for one, two and some multiview configurations. They provide practical recipes on how to photograph with RS cameras to avoid reconstruction errors. This method can be used to unblock some situations but it is not a solution to the standard SfM problem. Authors in [3] introduce 20pt and 44pt linear solvers for pure translational and uniform motion models respectively. However, the pure translational motion assumption is not feasible to model the camera motion in most of practical applications. Although more general, the 44-point solution requires too many correspondences and is therefore not suitable for use with RANSAC (Random Sample Consensus).

Parametrised homographies: Authors in [4] addressed the rectification and stabilization problem of RS videos by using a sequence of parametrised homographies (one for each image row pair from two consecutive frames). Camera poses are estimated for last

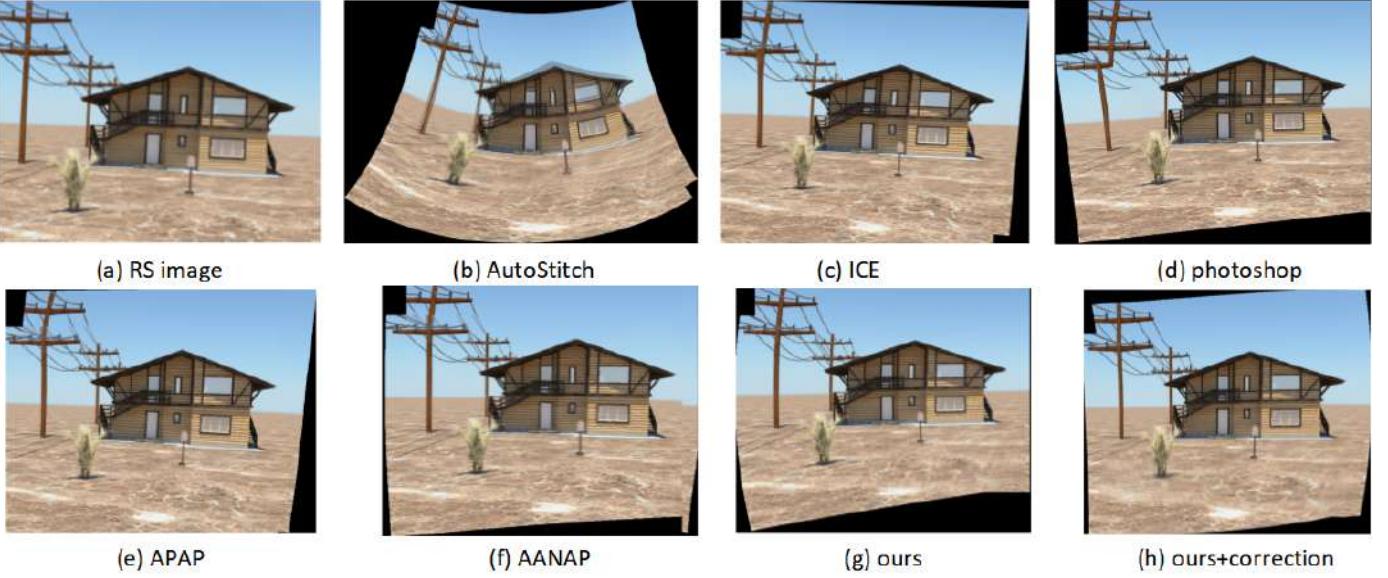


Fig. 2: RS images (a). Stitching results obtained with well-known commercial stitching applications such as AutoStitch [22] (b) Microsoft Image Composite Editor (ICE) [23] (c) Adobe Photoshop [24] (d) state-of-the-art multiple homographies stitching method APAP [25] (e) and AANAP [26] (f). The stitching results and the correction of the RS effects obtained with the the proposed method are shown in (g) and (h).

rows of each frame. Intermediate translations and rotations are interpolated from these. The camera motion is parametrised as a continuous curve. Curve parameters are solved for using non-linear least squares over inter-frame correspondences obtained by feature tracking. Thus, for general unordered set of images with large baselines, it is impossible for this method to enforce relationships between the camera poses and their motion (Fig. 1(b)). Besides, as in [16], pre-calibration of the inter-frame delay is required. Similarly, authors in [27] try to solve the RS correction problem by building multiple independent homographies between each of image rows. However, this method also requires multiple successive and smooth-motion frames as input since the use of a key-row interpolation approach [4]. Besides, only iterative nonlinear solution is provided for motion estimation.

One of the closest work to ours is [21], which employs a homography mixture and proposes a linear solution to estimate inter-frame motion. However, The major issue of this method is that points on a given row (row-block) in the first image have to be matched with points which also belong to a row (row-block) in the second image. This obviously limits the number of matches among unordered images even with small inter-frame motion.

In summary, as shown in Fig. 1, estimating relative poses from general unordered RS image sets remains an open problem and there is a need for new methods which require less input data (i.e. number of matches) and which work for various camera configurations. With some acceptable constraints on the scene structure or on the camera motion, homography could be used instead of epipolar geometry to recover the relative pose with less point matches [28], [29], [30]. It has many applications such as image registration and plane-based camera relative pose estimation. However, homography-based methods have not been fully studied and adapted to the RS case.

In this paper, we show that the relationship between points on a

RS pair can be expressed as a set of 3 to 7 atomic 3×3 matrices, depending on the kinematic model used for the instantaneous-motion during image acquisition. These matrices establishes a direct point-to-point transformation similarly to the homography matrix in the GS case. We call this set of matrices RS Homography. We also propose two linear solvers for the computation of this RS Homography using point correspondences. We first propose a theoretical 36pt linear solution and then derive a practical 13.5pt linear solver that gives good estimates of the homography between two RS views. We also investigate the use of the proposed method for two major computer vision applications:

(1) Plane-based Relative Pose Estimation: Although the RS relative pose problem has been addressed before [3], [16], a more efficient and robust solution which handles planar scenes is proposed by using RS homography in this paper.

(2) Image Stitching: Users nowadays frequently create panoramas by rotating RS cameras, e.g. ‘Pano’ mode in iPhone. Besides, 360 VR images such as Street View could also be taken with RS cameras installed on a moving platform such as the Google car [31]. As shown in Fig. 2(b-e), the most well-known commercial stitching software or state-of-the-art methods, which are based on the GS model, lead to poor results in presence of RS effects. Therefore, designing a RS stitching method by using RS Homography will be of valuable significance.

1.2 Paper Organization and Contributions

In this paper, we first introduce the RS instantaneous-motion and projection models in section 2. Then we derive the full and the simplified RS homography matrices in section 3 followed by the solutions to estimate these matrices from point matches in section 4. Next we show how to estimate the RS relative pose based on a planar scene in section 5 and also RS image stitching by using RS homography in section 6. Finally, we present experimental results of the proposed methods in section 8.

The contributions of this paper can be summarized as follows:

- This is the first work to investigate the theoretical RS homography matrices which can be expressed as a unique set of 3 to 8 atomic 3×3 matrices, depending on the kinematic model used for instantaneous-motion during image acquisition.
- We develop practical linear solvers for RS Homography matrices computation from two RS views with possibly large baseline (no constraint is done on the inter-frame motion, the kinematic model concerns only the instantaneous-motion).
- By using the RS Homography, we develop a linear solution for the plane-based relative pose estimation problem. We also introduce a solution to RS image stitching which not only can correctly align RS images (Fig. 2(g)) but also removes the RS effects at the same time (Fig. 2(h)).

2 ROLLING SHUTTER PROJECTION MODEL

RS Projection Model. Let a 3D point $\mathbf{P}_i = [x, y, z]^\top$ (w.r.t to the world coordinate system) be projected into a calibrated RS camera as an image point $\mathbf{q}_i = [u_i, v_i, 1]^\top$:

$$s_i \mathbf{K}^{-1} \mathbf{q}_i = \mathbf{R}_i \mathbf{P} + \mathbf{t}_i \quad (1)$$

where, s_i is the scene depth of the corresponding camera. $[\mathbf{R}_i, \mathbf{t}_i]$ is the camera pose when scanline (row) v_i is taken. \mathbf{K} is the camera calibration matrix. For the readability, image points \mathbf{q} are pre-multiplied by \mathbf{K}^{-1} in the rest of the paper.

We use the linearized uniform model to constrain the instantaneous-motion of a RS camera during the frame acquisition, which has been justified in previous works [3], [15], [16]. Thus, given the angular velocity $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^\top$ and the translational velocity $\mathbf{d} = [d_x, d_y, d_z]^\top$, we can express the camera pose $[\mathbf{R}_i | \mathbf{t}_i]$ at row v_i as:

$$\mathbf{R}_{v_i} = (\mathbf{I} + [\boldsymbol{\omega}]_{\times} v_i) \mathbf{R}_0 \quad \mathbf{t}_{v_i} = \mathbf{t}_0 + \mathbf{d} v_i \quad (2)$$

where \mathbf{I} is the 3×3 identity matrix, $[\boldsymbol{\omega}]_{\times}$ is a skew-symmetric matrix produced by $\boldsymbol{\omega}$, and $[\mathbf{R}_0 | \mathbf{t}_0]$ is the camera pose of the first row.

Rolling Shutter Relative Pose. Let us consider n 3D points $\mathbf{P}_i, i \in [1, n]$ imaged from two RS views with the camera poses $[\mathbf{R}_{v_i} | \mathbf{t}_{v_i}]$ and $[\mathbf{R}_{v'_i} | \mathbf{t}_{v'_i}]$ as $\mathbf{q}_i = [u_i, v_i, 1]^\top$ and $\mathbf{q}'_i = [u'_i, v'_i, 1]^\top$ in the two images respectively. $[\mathbf{I} | \mathbf{0}]$ and $[\mathbf{R}_0 | \mathbf{t}_0]$ are the camera poses of the first row of each image.

We assume that \mathbf{P}_i in world coordinates can also be expressed as \mathbf{P}_{v_i} and $\mathbf{P}_{v'_i}$ in the two camera coordinate systems. The transformations are written as:

$$\mathbf{P}_{v_i} = \mathbf{R}_{v_i} \mathbf{P} + \mathbf{t}_{v_i} \quad (3)$$

and

$$\mathbf{P}_{v'_i} = \mathbf{R}_{v'_i} \mathbf{P} + \mathbf{t}_{v'_i} \quad (4)$$

By substituting Eq. 3 into Eq. 4 and eliminating \mathbf{P} , we obtain the transformation between \mathbf{P}_{v_i} and $\mathbf{P}_{v'_i}$ as:

$$\begin{aligned} \mathbf{P}_{v'_i} &= \mathbf{R}_{v'_i} \mathbf{P} + \mathbf{t}_{v'_i} = \mathbf{R}_{v'_i} (\mathbf{R}_{v_i}^\top (\mathbf{P}_{v_i} - \mathbf{t}_{v_i})) + \mathbf{t}_{v'_i} \\ &= \underbrace{\mathbf{R}_{v'_i} \mathbf{R}_{v_i}^\top}_{\mathbf{R}_i} \mathbf{P}_{v_i} + \underbrace{\mathbf{t}_{v'_i} - \mathbf{R}_{v'_i} \mathbf{R}_{v_i}^\top \mathbf{t}_{v_i}}_{\mathbf{t}_i} \end{aligned} \quad (5)$$

Thus, the rotation \mathbf{R}_i and the translation \mathbf{t}_i between the the row v_i in the first image and the row v'_i in the second image are:

$$\begin{aligned} \mathbf{R}_i &= (\mathbf{I} + [\boldsymbol{\omega}_2]_{\times} v'_i) \mathbf{R}_0 (\mathbf{I} - [\boldsymbol{\omega}_1]_{\times} v_i) \\ \mathbf{t}_i &= \mathbf{t}_0 + \mathbf{d}_2 v'_i - (\mathbf{I} + [\boldsymbol{\omega}_2]_{\times} v'_i) \mathbf{R}_0 (\mathbf{I} - [\boldsymbol{\omega}_1]_{\times} v_i) \mathbf{d}_1 v_i \end{aligned} \quad (6)$$

where $\{\boldsymbol{\omega}_1, \mathbf{d}_1\}$ and $\{\boldsymbol{\omega}_2, \mathbf{d}_2\}$ are instantaneous-motion parameters of the two RS views.

3 RS HOMOGRAPHY

3.1 GS Homography Matrix

Let us assume that a planar object (a plane) is observed from two GS cameras with the poses $[\mathbf{I} | \mathbf{0}]$ and $[\mathbf{R}_0 | \mathbf{t}_0]$. The transformation between the two corresponding image points \mathbf{q}_i and \mathbf{q}'_i can be written as:

$$\alpha_i \mathbf{q}'_i = \mathbf{H}_{GS} \mathbf{q}_i = (\mathbf{R}_0 - \frac{\mathbf{t}_0 \mathbf{n}^\top}{d}) \mathbf{q}_i, \quad \alpha_i = z'/z \quad (7)$$

where α_i is a scale factor that depends on the depth of \mathbf{P}_i in each camera. \mathbf{H}_{GS} is the GS homography matrix, \mathbf{n} is the normal vector of the observed plane and d is the distance from the first camera to the plane under the constraint $\mathbf{n}^\top \mathbf{P}_i + d = 0$.

3.2 RS Homography Matrix

Expressions of the normal vector and the distance to plane in an RS frame. When instantaneous-motion occurs during the acquisition, \mathbf{H} between two RS cameras varies with different row combinations. The relative pose between the row v_i in the first image and the row v'_i in the second image is defined in Eq. (6). However, due to the instantaneous-motion of the first view, the normal vector of the plane and the distance to the plane are also changing dynamically with different scanlines (rows).

Firstly, we consider an RS camera is at the pose of its first row, then the plane constraint is:

$$\mathbf{n}_0^\top \mathbf{P}_i + d_0 = 0 \quad (8)$$

By substituting the transform from \mathbf{P}_i to \mathbf{P}_{v_i} in Eq. (3) into Eq. (8), we obtain:

$$\begin{aligned} &\mathbf{n}_0^\top (\mathbf{R}_{v_i}^\top (\mathbf{P}_{v_i} - \mathbf{t}_{v_i})) + d_0 \\ &= \underbrace{\mathbf{n}_0^\top \mathbf{R}_{v_i}^\top}_{\mathbf{n}_{v_i}^\top} \mathbf{P}_{v_i} + \underbrace{d_0 - \mathbf{n}_0^\top \mathbf{R}_{v_i}^\top \mathbf{t}_{v_i}}_{d_{v_i}} = 0 \end{aligned} \quad (9)$$

Then by substituting linear instantaneous-motion model Eq. (2) into Eq. (9), we can finally express the normal vector of the plane and distance to the plane w.r.t row v_i as:

$$\begin{aligned} \mathbf{n}_i^\top &= \mathbf{n}_0^\top (\mathbf{I} - [\boldsymbol{\omega}_1]_{\times} v_i) \\ d_i &= d_0 - \mathbf{n}_0^\top (\mathbf{I} - [\boldsymbol{\omega}_1]_{\times} v_i) \mathbf{d}_1 v_i \end{aligned} \quad (10)$$

where \mathbf{n}_0 and d_0 are the normal vector and the distance for the first row.

Derivation of RS Homography matrices. For convenience, we denote $\mathbf{R}_i, \mathbf{t}_i$ in Eq. (6) and \mathbf{n}_i^\top, d_i in Eq. (10) as:

$$\begin{aligned} \mathbf{R}_i &= \mathbf{R}_0 + \mathbf{R}_1 v_i + \mathbf{R}_2 v'_i + \mathbf{R}_3 v_i v'_i \\ \mathbf{t}_i &= \mathbf{t}_0 + \mathbf{t}_1 v_i + \mathbf{t}_2 v'_i + \mathbf{t}_3 v_i^2 + \mathbf{t}_4 v_i v'_i + \mathbf{t}_5 v_i^2 v'_i \\ -\frac{\mathbf{n}_i^\top}{d_i} &\approx -\frac{\mathbf{n}_0^\top - \mathbf{n}_0^\top [\omega_1] \times v_i}{d_0} = \mathbf{N}_0 + \mathbf{N}_1 v_i \end{aligned} \quad (11)$$

where, the auxiliary variables are defined as:

$$\left\{ \begin{array}{l} \mathbf{R}_0 = \mathbf{R}_0 \\ \mathbf{R}_1 = -\mathbf{R}_0 [\omega_1] \times \\ \mathbf{R}_2 = [\omega_2] \times \mathbf{R}_0 \\ \mathbf{R}_3 = -[\omega_2] \times \mathbf{R}_0 [\omega_1] \times \end{array} \right. \quad \left\{ \begin{array}{l} \mathbf{t}_0 = \mathbf{t}_0 \\ \mathbf{t}_1 = -\mathbf{R}_0 \mathbf{d}_1 \\ \mathbf{t}_2 = \mathbf{d}_2 \\ \mathbf{t}_3 = \mathbf{R}_0 [\omega_1] \times \mathbf{d}_1 \\ \mathbf{t}_4 = -[\omega_2] \times \mathbf{R}_0 \mathbf{d}_1 \\ \mathbf{t}_5 = [\omega_2] \times \mathbf{R}_0 [\omega_1] \times \mathbf{d}_1 \end{array} \right.$$

$$\left\{ \begin{array}{l} \mathbf{N}_0 = -\frac{\mathbf{n}_0^\top}{d_0} \\ \mathbf{N}_1 = \frac{\mathbf{n}_0^\top [\omega_1] \times}{d_0} \end{array} \right.$$

Then, by substituting Eq. (11) into Eq. (7), we can obtain:

$$\begin{aligned} \mathbf{H}_{RS,i} &= \mathbf{R}_i - \frac{\mathbf{t}_i \mathbf{n}_i^\top}{d_i} \\ &= (\mathbf{R}_0 + \mathbf{R}_1 v_i + \mathbf{R}_2 v'_i + \mathbf{R}_3 v_i v'_i) \\ &\quad + (\mathbf{t}_0 + \mathbf{t}_1 v_i + \mathbf{t}_2 v'_i + \mathbf{t}_3 v_i^2 + \mathbf{t}_4 v_i v'_i + \mathbf{t}_5 v_i^2 v'_i) \\ &= (\underbrace{\mathbf{R}_0 + \mathbf{t}_0 \mathbf{N}_0}_{\mathbf{H}_{GS}}) + (\underbrace{\mathbf{R}_1 + \mathbf{t}_1 \mathbf{N}_0 + \mathbf{t}_0 \mathbf{N}_1}_{\mathbf{H}_1}) v_i \\ &\quad + (\underbrace{\mathbf{R}_2 + \mathbf{t}_2 \mathbf{N}_0}_{\mathbf{H}_2}) v'_i + (\underbrace{\mathbf{R}_3 + \mathbf{t}_4 \mathbf{N}_0 + \mathbf{t}_2 \mathbf{N}_1}_{\mathbf{H}_3}) v_i v'_i \\ &\quad + (\underbrace{\mathbf{t}_3 \mathbf{N}_0 + \mathbf{t}_1 \mathbf{N}_1}_{\mathbf{H}_4}) v_i^2 + (\underbrace{\mathbf{t}_5 \mathbf{N}_0 + \mathbf{t}_4 \mathbf{N}_1}_{\mathbf{H}_5}) v_i^2 v'_i \\ &\quad + (\underbrace{\mathbf{t}_3 \mathbf{N}_1}_{\mathbf{H}_6}) v_i^3 + (\underbrace{\mathbf{t}_5 \mathbf{N}_1}_{\mathbf{H}_7}) v_i^3 v'_i \end{aligned} \quad (12)$$

thus, we can rewrite Eq. (12) in a simplified form by using eight 3×3 atomic matrices $\mathbf{H}_{GS}, \mathbf{H}_1 \dots \mathbf{H}_7$ as:

$$\mathbf{H}_{RS,i} = \mathbf{H}_{GS} + \mathbf{H}_1 v_i + \mathbf{H}_2 v'_i + \mathbf{H}_3 v_i v'_i + \mathbf{H}_4 v_i^2 + \mathbf{H}_5 v_i^2 v'_i + \mathbf{H}_6 v_i^3 + \mathbf{H}_7 v_i^3 v'_i \quad (13)$$

We draw attention to the fact that despite the presence of row indexes v_i and v'_i in the expression of \mathbf{H}_{RS} , matrices \mathbf{H}_{GS} and \mathbf{H}_1 to \mathbf{H}_7 do not depend on these indexes and are unique for all the point matches of any image pair.

3.3 Simplified RS Homography Matrix

Approximation of RS relative pose. Under the small rotation assumption, the second order and higher terms in Eq. (6) can be ignored. This simplification method is also used in [2], [17], [32], [33]. This approximation can be justified in that we force the translational speed vectors \mathbf{d}_1 and \mathbf{d}_2 to be constant in the world coordinate system, which is physically coherent with the constant velocity kinematic model. Therefore, we obtain an approximate expression of RS relative pose:

$$\begin{aligned} \mathbf{R}_i &= \mathbf{R}_0 - \mathbf{R}_0 [\omega_1] \times v_i + [\omega_2] \times \mathbf{R}_0 v'_i \\ \mathbf{t}_i &= \mathbf{t}_0 + \mathbf{d}_2 v'_i - \mathbf{R}_0 \mathbf{d}_1 v_i \end{aligned} \quad (14)$$

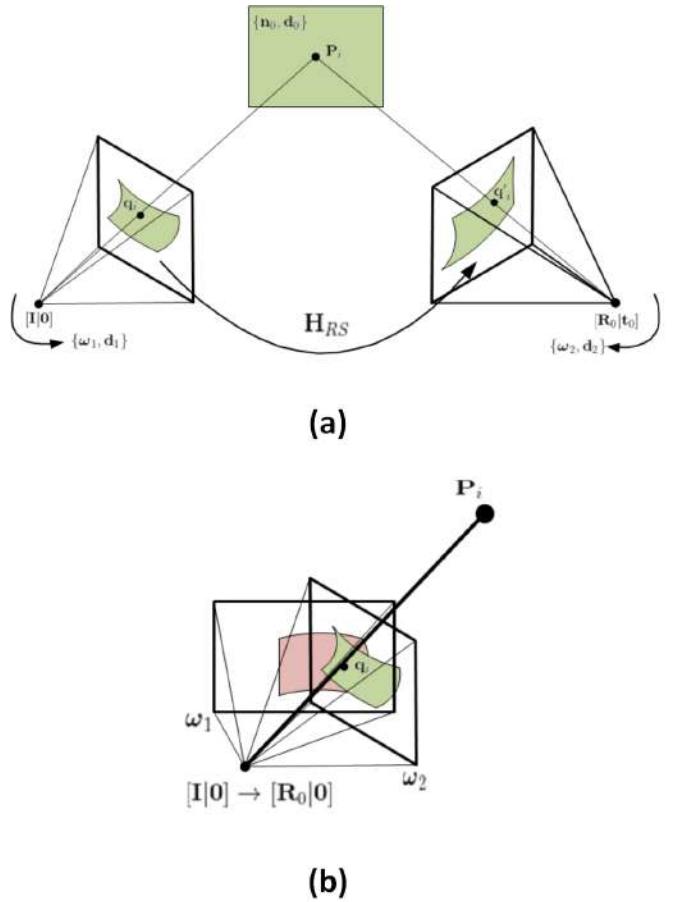


Fig. 3: Examples illustrating the general RS Homography case (a) and the rotate-only case (b).

Approximation of the plane pose. In practice, since the translation during acquisition is commonly much smaller than the distance from the camera to the scene plane, we can ignore the terms affected by translational velocities. In addition, we drop the second order terms, and obtain the approximate expressions:

$$\begin{aligned} \mathbf{n}_i^\top &= \mathbf{n}_0^\top (\mathbf{I} - [\omega_1] \times v_i) \\ d_i &= d_0 - \mathbf{n}_0^\top \mathbf{d}_1 v_i \approx d_0 \end{aligned} \quad (15)$$

Using both approximations in Eq. (14) and (15), the RS homography matrix \mathbf{H}_{RS} between the row v_i and the row v'_i in the two images can be simplified:

$$\begin{aligned} \mathbf{H}_{RS,i} &= \mathbf{H}_{GS} + \mathbf{A}_1 v_i + \mathbf{A}_2 v'_i \\ \text{where, } \mathbf{A}_1 &= -\mathbf{R}_0 [\omega_1] \times + \frac{\mathbf{R}_0 \mathbf{d}_1 \mathbf{n}_0^\top}{d_0} + \frac{\mathbf{t}_0 \mathbf{n}_0^\top [\omega_1] \times}{d_0} \\ \mathbf{A}_2 &= [\omega_2] \times \mathbf{R}_0 - \frac{\mathbf{d}_2 \mathbf{n}_0^\top}{d_0} \end{aligned} \quad (16)$$

where \mathbf{A}_1 and \mathbf{A}_2 are two atomic matrices. Note that RS Homography consists of the GS homography matrix \mathbf{H}_{GS} defined in Eq. (7) and of the two matrices $\mathbf{A}_1, \mathbf{A}_2$ which contain the instantaneous-motion parameters.

4 RS HOMOGRAPHY ESTIMATION

Now, we show how to linearly estimate the RS homographies defined in Eq. (13) and (16).

4.1 4pt GS Homography Estimation

The GS homography matrix is usually computed using the Direct Linear Transform algorithm (DLT). Based on Eq. (7), the cross product is given by $\mathbf{q}'_{v_i} \times \mathbf{H}_{GS}\mathbf{q}_i = \mathbf{0}$. This gives two linearly independent equations:

$$\begin{bmatrix} \mathbf{0}^\top & -\mathbf{q}_i^\top & v'_i \mathbf{q}_i^\top \\ \mathbf{q}_i^\top & \mathbf{0}^\top & -u'_i \mathbf{q}_i^\top \end{bmatrix} \begin{bmatrix} \mathbf{H}_{GS,(1)}^\top \\ \mathbf{H}_{GS,(2)}^\top \\ \mathbf{H}_{GS,(3)}^\top \end{bmatrix} = \mathbf{L}_i \mathbf{h}_{GS} = \mathbf{0} \quad (17)$$

where $\mathbf{0}^\top = [0, 0, 0]$. Given n point correspondences ($n \geq 4$), we obtain a system in the form $\mathbf{L}\mathbf{h}_{GS}$ where \mathbf{L} is a $2n \times 9$ matrix. The solution is then the singular vector associated to the smallest singular value of \mathbf{L} .

4.2 36pt Full RS Homography Matrix Estimation

In the RS case, the computation is less straightforward. By substituting Eq. (13) into \mathbf{h}_{GS} , Eq. (17) can be rewritten as follows:

$$\mathbf{M}_{RS,i} \mathbf{h}_{RS} = \begin{bmatrix} \mathbf{0} & \mathbf{q}_i \\ -\mathbf{q}_i & \mathbf{0} \\ v'_i \mathbf{q}_i & -u'_i \mathbf{q}_i \\ \mathbf{0} & v_i \mathbf{q}_i \\ -v_i \mathbf{q}_i & \mathbf{0} \\ v_i v'_i \mathbf{q}_i & -v_i u'_i \mathbf{q}_i \\ \mathbf{0} & v'_i \mathbf{q}_i \\ -v'_i \mathbf{q}_i & \mathbf{0} \\ v_i'^2 \mathbf{q}_i & -u'_i v'_i \mathbf{q}_i \\ \mathbf{0} & v_i v'_i \mathbf{q}_i \\ -v_i v'_i \mathbf{q}_i & \mathbf{0} \\ v_i v'_i^2 \mathbf{q}_i & -u'_i v_i v'_i \mathbf{q}_i \\ \mathbf{0} & v_i^2 \mathbf{q}_i \\ -v_i^2 \mathbf{q}_i & \mathbf{0} \\ v_i^2 v'_i \mathbf{q}_i & -u'_i v_i^2 \mathbf{q}_i \\ \mathbf{0} & v_i^2 v'_i \mathbf{q}_i \\ -v_i^2 v'_i \mathbf{q}_i & \mathbf{0} \\ v_i^2 v'_i^2 \mathbf{q}_i & -u'_i v_i^2 v'_i \mathbf{q}_i \\ \mathbf{0} & v_i^3 \mathbf{q}_i \\ -v_i^3 \mathbf{q}_i & \mathbf{0} \\ v_i^3 v'_i \mathbf{q}_i & -u'_i v_i^3 \mathbf{q}_i \\ \mathbf{0} & v_i^3 v'_i \mathbf{q}_i \\ -v_i^3 v'_i \mathbf{q}_i & \mathbf{0} \\ v_i^3 v'_i^2 \mathbf{q}_i & -u'_i v_i^3 v'_i \mathbf{q}_i \end{bmatrix}^\top \begin{bmatrix} \mathbf{H}_{GS,(1)}^\top \\ \mathbf{H}_{GS,(2)}^\top \\ \mathbf{H}_{GS,(3)}^\top \\ \mathbf{H}_{1,(1)}^\top \\ \mathbf{H}_{1,(2)}^\top \\ \mathbf{H}_{1,(3)}^\top \\ \mathbf{H}_{2,(1)}^\top \\ \mathbf{H}_{2,(2)}^\top \\ \mathbf{H}_{2,(3)}^\top \\ \mathbf{H}_{3,(1)}^\top \\ \mathbf{H}_{3,(2)}^\top \\ \mathbf{H}_{3,(3)}^\top \\ \mathbf{H}_{4,(1)}^\top \\ \mathbf{H}_{4,(2)}^\top \\ \mathbf{H}_{4,(3)}^\top \\ \mathbf{H}_{5,(1)}^\top \\ \mathbf{H}_{5,(2)}^\top \\ \mathbf{H}_{5,(3)}^\top \\ \mathbf{H}_{6,(1)}^\top \\ \mathbf{H}_{6,(2)}^\top \\ \mathbf{H}_{6,(3)}^\top \\ \mathbf{H}_{7,(1)}^\top \\ \mathbf{H}_{7,(2)}^\top \\ \mathbf{H}_{7,(3)}^\top \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (18)$$

where $\mathbf{H}_{(i)}$, $\mathbf{A}_{1,(i)}$ and $\mathbf{A}_{2,(i)}$ are the i^{th} rows of \mathbf{H}_{GS} , \mathbf{A}_1 and \mathbf{A}_2 respectively. $\mathbf{M}_{RS,i}$ is a 2×72 matrix, which consists of the terms of u_i , v_i and \mathbf{q}_i . \mathbf{h}_{RS} is a 72×1 vector that contains the components of \mathbf{H}_{GS} , $\mathbf{H}_1 \dots \mathbf{H}_7$.

Each point correspondence gives two constraints such as Eq. (18). Given n point correspondences ($n \geq 36$), we obtain a system in the form $\mathbf{M}_{RS} \mathbf{h}_{RS} = [0, \dots, 0]^\top$ where \mathbf{M} is a $2n \times 72$ matrix. This system is solved using SVD.

4.3 13.5pt Simplified RS Homography Estimation

In the simplified case of section 3.3, by substituting Eq. (16) into \mathbf{h}_{GS} , Eq. (17) can be rewritten as:

$$\begin{bmatrix} \mathbf{0} & \mathbf{q}_i \\ -\mathbf{q}_i & \mathbf{0} \\ v'_i \mathbf{q}_i & -u'_i \mathbf{q}_i \\ \mathbf{0} & v_i \mathbf{q}_i \\ -v_i \mathbf{q}_i & \mathbf{0} \\ v_i v'_i \mathbf{q}_i & -v_i u'_i \mathbf{q}_i \\ \mathbf{0} & v'_i \mathbf{q}_i \\ -v'_i \mathbf{q}_i & \mathbf{0} \\ v_i'^2 \mathbf{q}_i & -u'_i v'_i \mathbf{q}_i \end{bmatrix}^\top \begin{bmatrix} \mathbf{H}_{GS,(1)}^\top \\ \mathbf{H}_{GS,(2)}^\top \\ \mathbf{H}_{GS,(3)}^\top \\ \mathbf{A}_{1,(1)}^\top \\ \mathbf{A}_{1,(2)}^\top \\ \mathbf{A}_{1,(3)}^\top \\ \mathbf{A}_{2,(1)}^\top \\ \mathbf{A}_{2,(2)}^\top \\ \mathbf{A}_{2,(3)}^\top \end{bmatrix} = \mathbf{M}_{RS,i} \mathbf{h}_{RS} = [0 \ 0]^\top \quad (19)$$

where $\mathbf{M}_{RS,i}$ reduces to a 2×27 matrix and $\mathbf{h}_{RS,i}$ is a 27×1 vector with 27 unknowns. Thus, with at least 14 point correspondences, we can estimate \mathbf{h}_{RS} linearly by using SVD. In order to obtain stable results, we perform a normalization of \mathbf{M}_{RS} in the way explained in [3], [28].

We extend the standard RANSAC pipeline [34] for homography verification [22] with the proposed 13.5pt linear solver to obtain a robust estimate of the RS homography matrix and matching inliers.

5 PLANE-BASED RS RELATIVE POSE AND INSTANTANEOUS-MOTIONS ESTIMATION

The 36pt solver requires a large amount of point matches and therefore it is not a feasible solution for RANSAC. However, using the simplified RS homography brings inaccuracy.

Thus, we suggest the following pipeline: **1)** Solve the RS homography matrix by estimating the atomic matrices (\mathbf{H}_{GS} , \mathbf{A}_1 and \mathbf{A}_2) with the 13.5pt solver. **2)** Linearly extract the relative pose and instantaneous-motions. **3)** Refine all the parameters in a nonlinear optimization based on the full RS Homography.

5.1 Computing Relative Pose and Instantaneous-motions

Relative pose $[\mathbf{R}_0 | \mathbf{t}_0]$ and **plane normal vector** \mathbf{n}_0 . Once \mathbf{H}_{GS} is known, it can be decomposed into \mathbf{R}_0 , \mathbf{t}_0 and \mathbf{n}_0 by using SVD. \mathbf{d}_0 is set as 1 and absorbed by \mathbf{t}_0 . Generally, this decomposition yields four solutions, where only at most two are physically valid. The positive depth constraint can be used to find the final solution [35], [36].

Instantaneous-motions. We can further retrieve instantaneous-motion parameters thanks to two linear equation systems derived from matrices \mathbf{A}_1 and \mathbf{A}_2 :

(1) First we compute $\boldsymbol{\omega}_1 = \{\omega_1^x, \omega_1^y, \omega_1^z\}$ and $\mathbf{d}_1 = \{d_1^x, d_1^y, d_1^z\}$ by using the previously computed values of $[\mathbf{R}_0 | \mathbf{t}_0]$ and \mathbf{n}_0 in matrix \mathbf{A}_1 (6 unknowns with 9 equations): Based on the definition of \mathbf{A}_1 in Eq. (16), we obtain the following linear system in $\boldsymbol{\omega}_1$ and \mathbf{d}_1 (detailed derivation in supplemental materials):

$$\begin{bmatrix} 0 & -G_{13} & G_{12} & n_0^x \mathbf{R}_{0,(1)} \\ G_{13} & 0 & -G_{11} & n_0^y \mathbf{R}_{0,(1)} \\ -G_{12} & G_{11} & 0 & n_0^z \mathbf{R}_{0,(1)} \\ 0 & -G_{23} & G_{22} & n_0^x \mathbf{R}_{0,(2)} \\ G_{23} & 0 & -G_{21} & n_0^y \mathbf{R}_{0,(2)} \\ -G_{22} & G_{21} & 0 & n_0^z \mathbf{R}_{0,(2)} \\ 0 & -G_{33} & G_{32} & n_0^x \mathbf{R}_{0,(3)} \\ G_{33} & 0 & -G_{31} & n_0^y \mathbf{R}_{0,(3)} \\ -G_{32} & G_{31} & 0 & n_0^z \mathbf{R}_{0,(3)} \end{bmatrix} \begin{pmatrix} \omega_1^x \\ \omega_1^y \\ \omega_1^z \\ d_1^x \\ d_1^y \\ d_1^z \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{1,11} \\ \mathbf{A}_{1,12} \\ \mathbf{A}_{1,13} \\ \mathbf{A}_{1,21} \\ \mathbf{A}_{1,22} \\ \mathbf{A}_{1,23} \\ \mathbf{A}_{1,31} \\ \mathbf{A}_{1,32} \\ \mathbf{A}_{1,33} \end{pmatrix} \quad (20)$$

where the auxiliary matrix \mathbf{G} is defined as $\mathbf{G} = \mathbf{R}_{0,(i)} + \mathbf{t}_0 \mathbf{n}_0^\top$ and $\mathbf{R}_{0,(i)}$ is the i^{th} row of \mathbf{R}_0 . As a result, 6 unknowns in $\boldsymbol{\omega}_1$ and \mathbf{d}_1 can be obtained by solving Eq. (20) linearly.

(2) Then we extract $\boldsymbol{\omega}_2 = \{\omega_2^x, \omega_2^y, \omega_2^z\}$ and $\mathbf{d}_2 = \{d_2^x, d_2^y, d_2^z\}$ from \mathbf{A}_2 (6 unknowns with 9 equations): Based on the definition of \mathbf{A}_2 in Eq. (16), we obtain the following linear system in $\boldsymbol{\omega}_2$ and \mathbf{d}_2 (detailed derivation in supplemental materials):

$$\begin{bmatrix} \mathbf{0} & -\mathbf{R}_{0,(3)} & \mathbf{R}_{0,(2)} \\ \mathbf{R}_{0,(3)} & \mathbf{0} & \mathbf{R}_{0,(1)} \\ -\mathbf{R}_{0,(2)} & \mathbf{R}_{0,(1)} & \mathbf{0} \\ -\mathbf{n}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{n}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{n}_0 \end{bmatrix}^\top \begin{pmatrix} \omega_2^x \\ \omega_2^y \\ \omega_2^z \\ d_2^x \\ d_2^y \\ d_2^z \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{2,11} \\ \mathbf{A}_{2,12} \\ \mathbf{A}_{2,13} \\ \mathbf{A}_{2,21} \\ \mathbf{A}_{2,22} \\ \mathbf{A}_{2,23} \\ \mathbf{A}_{2,31} \\ \mathbf{A}_{2,32} \\ \mathbf{A}_{2,33} \end{pmatrix} \quad (21)$$

Thus $\boldsymbol{\omega}_2$ and \mathbf{d}_2 can be obtained by solving Eq. (21) linearly.

5.2 Nonlinear Refinement.

The final step consists in a nonlinear refinement of pose and instantaneous-motion parameters with n pairs of point matches which are the inliers from the 13.5pt-RANSAC. This is achieved by minimizing the following cost function where the full homography matrix is used:

$$\arg \min_{\mathbf{R}_0, \mathbf{t}_0, \mathbf{n}_0, d_0, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \mathbf{d}_1, \mathbf{d}_2} = \sum_{i=1}^n (\mathbf{M}_{RS,i} \mathbf{h}_{RS})^2 \quad (22)$$

where $\mathbf{M}_{RS,i}$ and \mathbf{h}_{RS} are defined in Eq. (18).

6 RS IMAGE STITCHING

6.1 Working Assumptions

The goal of image stitching is to create a very wide angle image (or a panorama) from a set of images. After finding the homography matrix that aligns each pair of neighboring cameras, all the images are transformed so that they are mapped into the same projective space.

For that purpose, the cameras are assumed to have rotated about (approximately) the same center of projection. Thus, the RS homography matrix is further simplified by setting \mathbf{t}_i , \mathbf{d}_1 and \mathbf{d}_2 to 0, which leads to $\mathbf{H}_{GS} = \mathbf{R}_0$, $\mathbf{A}_1 = -\mathbf{R}_0 [\boldsymbol{\omega}_1]_\times$ and $\mathbf{A}_2 = [\boldsymbol{\omega}_2]_\times \mathbf{R}_0$.

6.2 Stitching Pipeline

6.2.1 RS homography estimation with 13.5pt solver

The RS homography matrix of coinciding optical centres has the same structure as the simplified RS homography matrix in Eq. (16). Thus, the 13.5pt method (section 4.3) is also feasible here.

6.2.2 RS image alignment

When aligning two GS images all image points can be directly mapped to new locations by applying $\mathbf{q}'_i = \mathbf{H}_{GS} \mathbf{q}_i$. For two RS images, Eq. (16) shows that any pair of feature correspondence, a 2D homography relationship (\mathbf{H}_{RS}) however still holds locally by the controls of row indexes in the two images. However, in stitching case, we want to map all pixels in the first image to the second image without prior about the row indexes of their correspondents. Thus, based on Eq. (16), we will develop a **global** mapping function which align the two images.

Global RS homography mapping. Notice that the row index v'_i is present in both sides of Eq. (13). However, by transformation and substitution, the local homography relationship becomes a globally parametrized non-linear mapping $\eta(\mathbf{q}_i)$ defined as follows (detailed derivation in supplemental materials):

$$[u'_i \ v'_i]^\top = \eta(u_i, v_i) = \begin{cases} \alpha(u_i, v_i) \\ \beta(u_i, v_i) \end{cases}$$

where,

$$\alpha(u_i, v_i) = \frac{d}{e} \quad \text{and} \quad \beta(u_i, v_i) = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

with,

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{2,(3)}^\top & \mathbf{H}_{GS,(3)}^\top - \mathbf{H}_{GS,(2)}^\top & -\mathbf{H}_{GS,(2)}^\top \\ \mathbf{H}_{3,(3)}^\top & \mathbf{H}_{1,(3)}^\top - \mathbf{H}_{2,(2)}^\top & -\mathbf{H}_{1,(2)}^\top \\ \mathbf{H}_{5,(3)}^\top & \mathbf{H}_{4,(3)}^\top - \mathbf{H}_{5,(2)}^\top & -\mathbf{H}_{4,(2)}^\top \\ \mathbf{H}_{7,(3)}^\top & \mathbf{H}_{6,(3)}^\top - \mathbf{H}_{7,(2)}^\top & -\mathbf{H}_{6,(2)}^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{q}_i \\ \mathbf{q}_i v_i \\ \mathbf{q}_i v_i^2 \\ \mathbf{q}_i v_i^3 \end{bmatrix}$$

$$\begin{bmatrix} d \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{GS,(1)}^\top & \mathbf{H}_{GS,(3)}^\top \\ \mathbf{H}_{1,(1)}^\top v_i & \mathbf{H}_{1,(3)}^\top v_i \\ \mathbf{H}_{2,(1)}^\top \beta(u_i, v_i) & \mathbf{H}_{2,(3)}^\top \beta(u_i, v_i) \\ \mathbf{H}_{3,(1)}^\top v_i \beta(u_i, v_i) & \mathbf{H}_{3,(3)}^\top v_i \beta(u_i, v_i) \\ \mathbf{H}_{4,(1)}^\top v_i^2 & \mathbf{H}_{4,(3)}^\top v_i^2 \\ \mathbf{H}_{5,(1)}^\top v_i^2 \beta(u_i, v_i) & \mathbf{H}_{5,(3)}^\top v_i^2 \beta(u_i, v_i) \\ \mathbf{H}_{6,(1)}^\top v_i^3 & \mathbf{H}_{6,(3)}^\top v_i^3 \\ \mathbf{H}_{7,(1)}^\top v_i^3 \beta(u_i, v_i) & \mathbf{H}_{7,(3)}^\top v_i^3 \beta(u_i, v_i) \end{bmatrix}^\top \begin{bmatrix} \mathbf{q}_i \\ \mathbf{q}_i \\ \vdots \\ \mathbf{q}_i \end{bmatrix} \quad (23)$$

where $\mathbf{H}_{GS,(i)}$ is the i^{th} row of \mathbf{H}_{GS} . There are two feasible solutions for the mapping but only one is correct in practice. The strategy of selecting the correct solution will be discussed later.

Global RS homography mapping for simplified case. The Full RS homography matrix Eq. (13) yields global mapping Eq. (23). Similarly, simplified RS homography matrix Eq. (16) yields a new globally non-linear mapping and thus updates $\eta(\mathbf{q}_i)$ as follows (detailed derivation in supplemental materials):

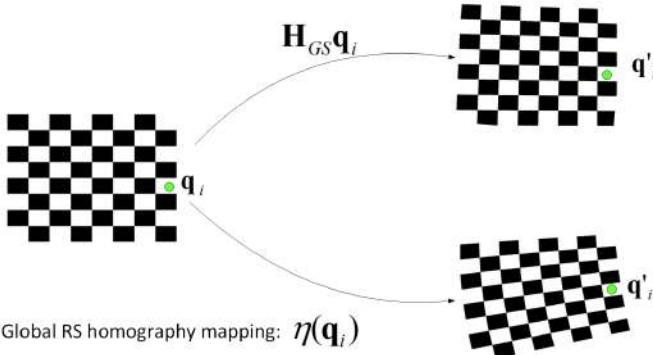


Fig. 4: Comparison of GS homography mapping and RS homography mapping.

$$[u'_i \ v'_i]^\top = \eta(u_i, v_i) = \begin{cases} \alpha(u_i, v_i) \\ \beta(u_i, v_i) \end{cases}$$

where,

$$\alpha(u_i, v_i) = \frac{d}{e} \quad \text{and} \quad \beta(u_i, v_i) = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

with,

$$\begin{aligned} a &= \mathbf{A}_{2,(3)} q_i \\ b &= \mathbf{H}_{GS,(3)} q_i + \mathbf{A}_{1,(3)} q_i v_i - \mathbf{A}_{2,(2)} q_i \\ c &= -\mathbf{H}_{GS,(2)} q_i - \mathbf{A}_{1,(2)} q_i v_i \\ d &= (\mathbf{H}_{GS,(1)} + \mathbf{A}_{1,(1)} v_i + \mathbf{A}_{2,(1)} \beta(u_i, v_i)) \mathbf{q}_i \\ e &= (\mathbf{H}_{GS,(3)} + \mathbf{A}_{1,(3)} v_i + \mathbf{A}_{2,(3)} \beta(u_i, v_i)) \mathbf{q}_i \end{aligned}$$

Similar to Eq. (23), it also leads to two feasible solutions.

It is important to realize that previous works [4], [21] use multiple **local** homographies, that are controlled by both row indexes in the two images. However, as shown in Fig. 4, η is determined only by the shared parameters \mathbf{H}_{GS} , $\mathbf{H}_1 \dots \mathbf{H}_7$ for all point matches. Thus, it describes a **global** and **unique** 2D mapping from q_i to q'_i directly.

Note that η is a nonlinear 2D mapping while the classical GS homography \mathbf{H}_{GS} is a linear one. But since η plays the same role as \mathbf{H}_{GS} , thus, we keep calling η as RS Homography mapping.

Since we choose 13.5pt solver to estimate the RS Homography, thus, for consistency, global RS homography mapping for simplified case Eq. (24) will be used for image alignment.

6.2.3 Selecting the Correct Solution

In Eq. (24), there are two feasible solutions for each pixel in original image. The solution with the smaller Euclidean distance to original point before warping is chosen as the correct one. This is usually the solution which maintains the consistency of the image registration. We will show through experiments that this is indeed effective.

6.2.4 Blending

To seamlessly blend the images, a multi-band blending strategy [37], [38] is used.

6.2.5 Correction of Stitched RS Image

After determining the instantaneous-motion parameters by means of the method described in section 5, an inverse mapping is applied to the aligned image points in order to remove RS distortions by compensating camera instantaneous-motion as follows [7]:

$$\mathbf{q}^{correct} = \mathbf{K}_2(\mathbf{I} - [\omega_2]_\times)\mathbf{q}' \quad (25)$$

Note that we correct the RS image after stitching. Another solution is perform the pre-rectification to the both RS images after RS homography estimation using ω_1 and ω_2 respectively, followed by classical GS image stitching. In our experiments, we found the choice of 'stitching first' to work better than the other and therefore present only those results.

6.3 Uncalibrated Image Stitching

Direct RS image alignment: In image stitching applications, it is common that the input images are uncalibrated. Considering that points \mathbf{q}_i and \mathbf{q}'_i are the image measurements in pixels (i.e. $[u_i, v_i, 1]$) instead of normalized points, Eq. (19) gives the uncalibrated $\mathbf{H}_{RS,i}^{image}$ which is defined as follows: $\mathbf{H}_{GS}^{image} = \mathbf{K}_2 \mathbf{H}_{GS} \mathbf{K}_1^{-1}$, $\mathbf{A}_1^{image} = \mathbf{K}_2 \mathbf{A}_1 \mathbf{K}_1^{-1}$ and $\mathbf{A}_2^{image} = \mathbf{K}_2 \mathbf{A}_2 \mathbf{K}_1^{-1}$. \mathbf{K}_1 and \mathbf{K}_2 are the calibration matrices of the first and second camera respectively.

Thus if normalized image points are replaced by the image measurements in Eq. (19) and Eq. (24), we can estimate $\mathbf{H}_{RS,i}^{image}$ by using the 13.5pt method (section 4.3) and align the images directly in the image space. This means that we can stitch two RS images without prior knowledge of the calibration matrices \mathbf{K}_1 and \mathbf{K}_2 .

RS correction with non-linear refinement: However, the determination of the instantaneous-motion parameters from \mathbf{A}_1^{image} and \mathbf{A}_2^{image} is different from the decomposition method of \mathbf{A}_1 and \mathbf{A}_2 described in section 5 since the calibration matrices are unknown.

Thus, based on the pin-hole camera model, we assume that principle point is located in the centre of the image. Thus only the focal length f remains unknown. Now, the problem is to estimate the focal length f and instantaneous-motions ω_1 and ω_2 given \mathbf{H}_{GS}^{image} , \mathbf{A}_1^{image} and \mathbf{A}_2^{image} . We first set the focal length as 0.9 times of the maximal dimension of each corresponding image [33]. By using the direct relative pose and instantaneous-motion algorithm in section 5, we can roughly estimate the rotation between the two images and angular velocities. Finally, we perform an iterative refinement to estimate the focal lengths, the rotation between cameras and instantaneous-motions as follows:

$$\arg \min_{f_1, f_2, \mathbf{R}_0, \mathbf{n}_0, \omega_1, \omega_2} = \sum_{i=1}^n (\mathbf{M}_{RS,i} \mathbf{h}_{RS} - \mathbf{b}_i)^2 \quad (26)$$

With the estimated parameters, an inverse mapping similar to Eq. (25) is applied to the stitched image directly, as follows:

$$\mathbf{q}^{correct} = \mathbf{K}_2(\mathbf{I} - [\omega_2]_\times)\mathbf{K}_2^{-1}\mathbf{q}' \quad (27)$$

7 DISCUSSIONS

The strategy which consists in combining 13.5pt and 36pt solvers offers an efficient and numerically stable solution for the RS Homography estimation problem. It enables us to compute the full set of matrices that play, in the case of RS images, the same

role than the classical GS homography matrix. Relative pose and instantaneous motion of both cameras can then be retrieved by decomposing these matrices without a need for any initial guess to be provided by users.

Nevertheless, the proposed solver is not a minimal one since it needs more correspondences than specified by the chosen formalism based on constant velocity model which is generally adopted in the literature being a good compromise between simplicity and precision. Indeed, 13.5 point correspondences are used instead of 10.

In the last decade, methods based on Grobner bases have been used successfully to solve several minimal problems in computer vision [39]. This was favored by the use of automatic generators of polynomial solvers such as [40] and more recently [41]. However, as the number of variables and equations grows the process may become intractable in practice. Although we were not able to generate a convenient solver directly from our equation system using [40], a more handcrafted system may potentially lead to a more minimal solver.

8 EXPERIMENTS

Both the RS homography-based pose estimation (RSH) and image stitching method presented in this paper were evaluated on synthetic and real data.

8.1 Relative Pose Estimation

8.1.1 Synthetic Data Experiments

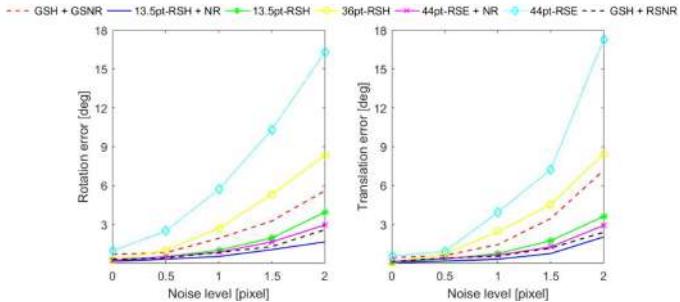


Fig. 5: Errors of relative pose estimation with increasing image noise.

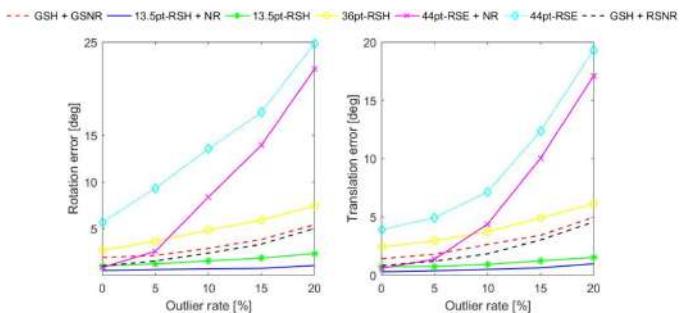


Fig. 6: Errors of relative pose estimation with increasing outlier rate.

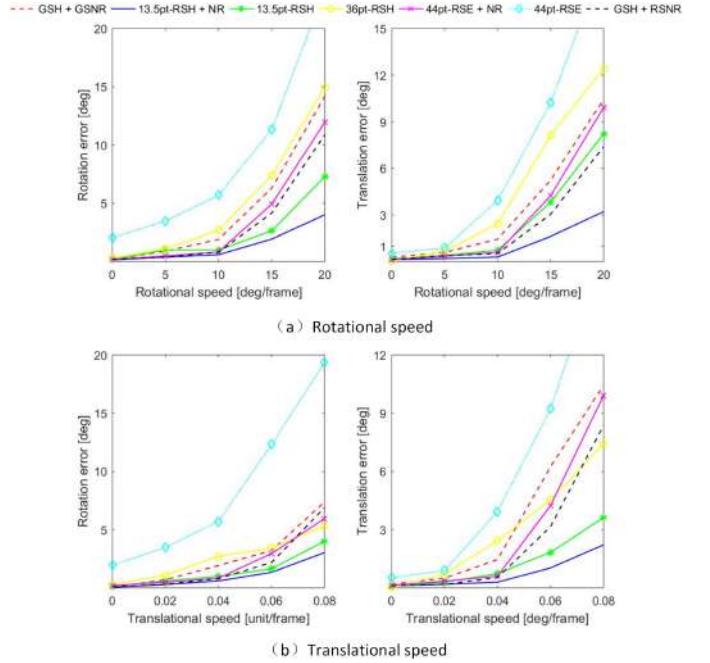


Fig. 7: Errors of relative pose estimation with increasing rotational speed (a) and translational speed (b).

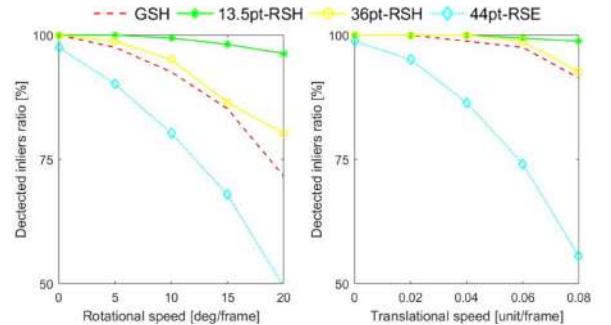


Fig. 8: Rate of inliers with increasing rotational speed and translational speed.

Experiment settings. We generated a plane scene with 60 feature points, which was imaged by two RS cameras with 480×640 image resolution. We set the distance from the plane to the optical centre of the first camera as 1 unit, and located the second camera randomly on a sphere around the centre of the plane with 1 unit length radius.

Since the ground truth of the relative pose is known, we calculated the relative pose error as follow:

- Rotation error: $e_{\text{rot}} = \arccos((\text{tr}(\mathbf{R}\mathbf{R}_{\text{GT}}^{\top}) - 1)/2)$
- Translation error: $e_{\text{trans}} = \arccos(\mathbf{t}^{\top}\mathbf{t}_{\text{GT}}/(\|\mathbf{t}\| \|\mathbf{t}_{\text{GT}}\|))$

The results are obtained after averaging the errors over 50 trials (the default setting is 1p noise, 0% outlier rate, 10 degs/frame and 0.04 units/frame for the rotational and translational speed).

Competing methods. We denote the proposed linear RS homography solvers using simplified matrix model with RANSAC verification as **13.5pt-RSH** and full matrix model as **36pt-RSH**. The nonlinear refinement described in Eq. (22) is denoted as **NR**.

We compared our method with two other methods:

- **GSH:** Classical GS homography estimation with RANSAC and decomposition¹.
- **GSH + GSNR:** GS homography RANSAC followed by nonlinear refinement using GS homography model **GSNR** (GS plane-based bundle adjustment described but with known calibration matrix [29]).
- **GSH + RSNR:** GS homography RANSAC followed by nonlinear refinement using GS homography model **RSNR** (Eq. (22)).
- **44pt-RSE:** RS essential matrix estimation with RANSAC and decomposition [3].
- **44pt-RSE + NR:** RS essential matrix estimation with RANSAC followed by nonlinear refinement **NR** (minimizing Sampson error).

(1) Stability vs Pixel noise: We first tested the stability of the proposed method in the presence of image noise. We increased the random Gaussian image noise from 0 to $2p$. Results in Fig. 5 show that both **36pt-RSH** and **44pt-RSE** present numerical instabilities with increasing noise level. In contrast, **13.5pt-RSH** provides stable and accurate estimations. Note that the performance of **44pt-RSE** is significantly improved after performing **NR** but remains worse than **36pt-RSH + NR**.

(2) Stability vs Outlier rate: We also tested the impact of outlier rate on the stability of the GS-based method, the epipolar-geometry-based method and the proposed method. The outlier correspondences are created by setting the coordinates of these outliers randomly within the range of RS images. As shown in Fig. 6, with increasing outlier rate, **44pt-RSE** fails in finding the correct inlier matches and provides wrong estimations which even can not be refined by performing **NR**. Similarly, neither of **GSH + GSNR** and **GSH + RSNR** can refine the rough estimations provided by **GSH**. We interpret this observation as **GSH** fails in filtering outliers. Conversely, the proposed **13.5pt-RSH** method obtains significantly more accurate results and shows good outlier rejection performance in RANSAC.

(3) Accuracy vs Instantaneous velocity: We also evaluated the performances by increasing the rotational speed from 0 to 20 degs/frame and the translational speed from 0 to 0.08 units/frame respectively. As shown in Fig. 7, **44pt-RSE + NR**, **GSH + GSNR** and **GSH + RSNR** provide large errors with the increasing instantaneous speed. In contrast, our method obtains obvious improvements comparing to all the competing methods.

Besides, we investigated the influence of the RS instantaneous-motions on RANSAC-based determination of inliers. As shown in Fig.8, with the increase of the camera instantaneous-motion, the inlier detection rate of **GSH**, **44pt-RSE** and **36pt-RSH** decreases dramatically. In contrast, the proposed **13.5pt-RSH** maintains its good performance in the determination.

8.1.2 Real Data Experiment

For readability, in the following sections, we denote the 13.5pt RS homography method with RANSAC and non-linear refinement as **RSH** (camera poses are final refine by RS bundle adjustment [17]) and GS-based homography with RANSAC and non-linear refinement [29] as **GSH**.

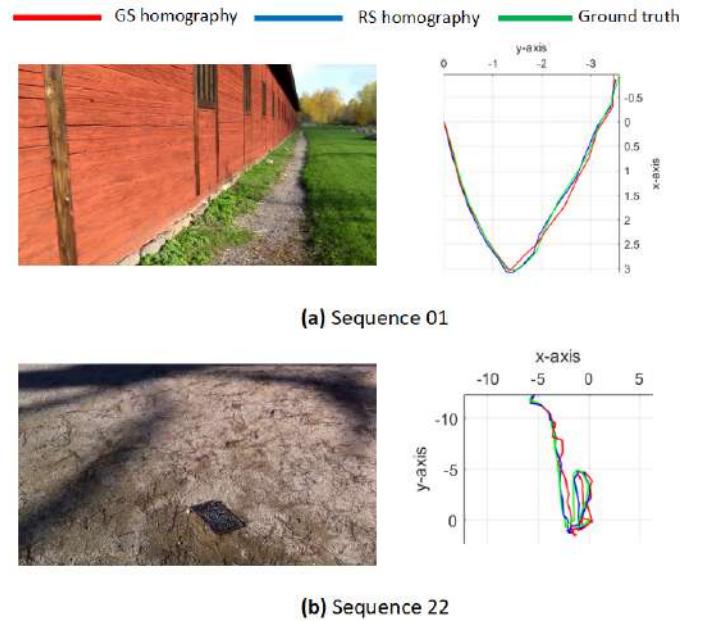


Fig. 9: Comparison of trajectory estimation (right sides) by using **GSH** and **RSH** on two RS image sequences (examples of input RS images are shown on the left side).

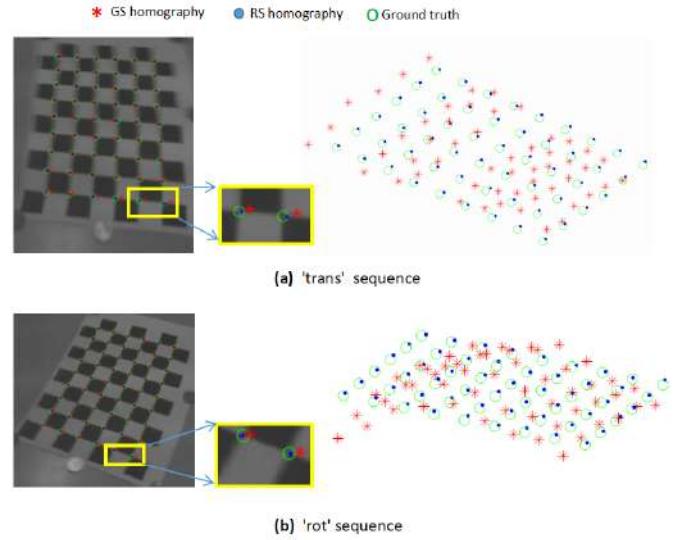


Fig. 10: Mapping and 3D reconstruction errors by using **GSH** and **RSH** on sequence 'trans' and 'rot' respectively.

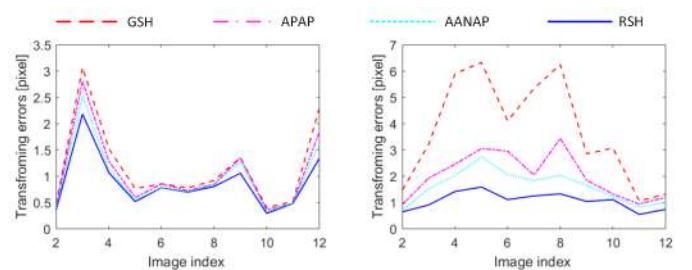


Fig. 11: Mapping errors by using **GSH**, **APAP**, **AANAP** and **RSH** on sequence 'rot1' and 'rot2' respectively.

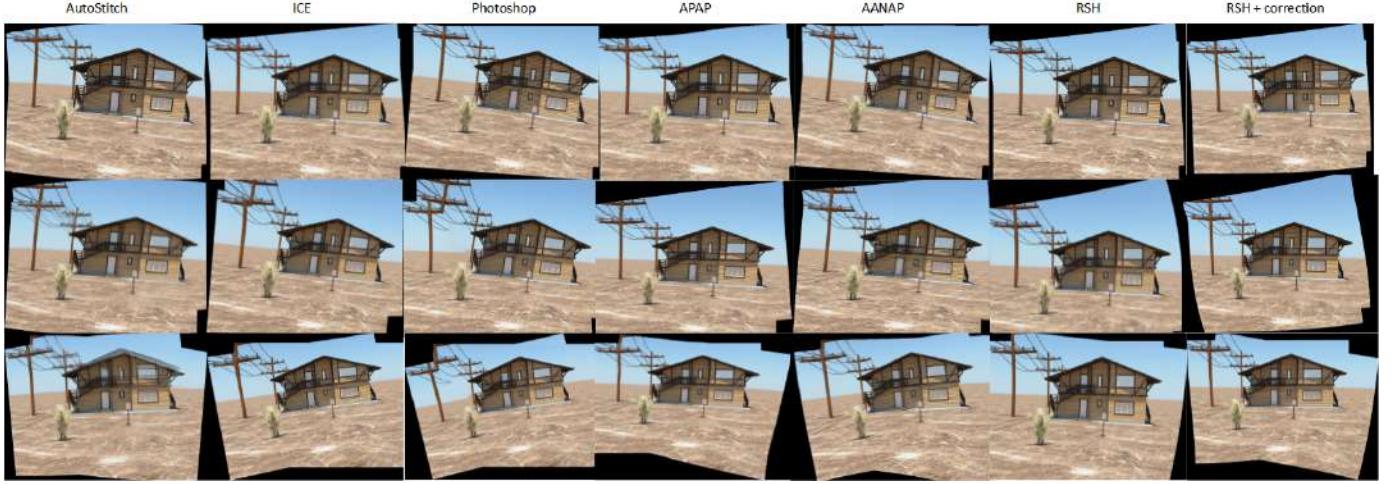


Fig. 12: Results of synthetic RS images stitching with different methods.

TABLE 1: Error of the poses estimated from rolling shutter sequences against to ground truth.

sequence	seq01 [unit]	seq22 [unit]m
GSH	0.903	3.620
RSH	0.199	1.668

TABLE 2: Errors of the trajectory estimation, 2D transformation and 3D Reconstruction

seq		GSH	RSH
'trans'	Mapping errors [pixel]	2.824	1.162
	Reconstruction errors [unit]	1.833	0.317
'rot'	Mapping errors [pixel]	4.131	2.005
	Reconstruction errors [unit]	2.519	0.397

(1) Plane-based trajectory estimation. For this experiment we used sequence '01' and '22' from [14] which was captured by a camera rig consisted of an iPhone4 camera (RS) and a Canon S95 camera (GS) simultaneously. However, contrarily to [14] and [16] which require smooth instantaneous-motion input video, the proposed method can handle large baselines. Thus we just selected non-successive frames with 9 frames interval as an input sequence. Since the ground truth can be computed by using the corresponding frames from GS sequence, we used the method described in [14] to calculate the pose error. The visual and quantitative evaluations summarized in Fig. 9 and Table 1 show that the proposed **RSH** method performs significantly better than **GSH** in both sequences.

(2) Plane-based SfM. We evaluated both **GSH** and **RSH** on two more challenging RS image sequences from [10]: 'trans' and 'rot' which are taken with mainly translational and rotational velocities respectively. Two RS images with the integrated visible chessboard are chosen from the same sequence to estimate the relative pose with **GSH** and **RSH**. Then we perform a triangulation to reconstruct the chessboard (note that the pose of a row of RS image is obtained by using Eq. (2)). Since the ground-truth of the

poses are unknown, we do the quantitative comparisons using the following two methods:

- 1) Mapping error: average homography mapping errors of the feature points in the chessboard from the first image to the second. For each point correspondence, the mapping error can be computed as: $e_i = \|\mathbf{q}'_i - \tau(\mathbf{q}_i)\|$, where τ is the 2D homography mapping. Thus, we set $\tau(\mathbf{q}_i) = \mathbf{H}_{GSH}\mathbf{q}_i$ for **GSH** while $\tau(\mathbf{q}_i) = \eta(\mathbf{q}_i)$ for **RSH**.
- 2) The reconstruction errors of the chessboard (each reconstructed 3D point is spatially aligned with the ground-truth, by minimizing the sum of all squared point-to-point distances).

The results are presented in Fig. 10 and Table 2 and show that **RSH** obtains significantly better results compared to **GSH** in both sequences.

8.2 Image stitching

8.2.1 Compared Methods

The image stitching experiments were conducted under the following settings:

- 1) The input images are **uncalibrated** (calibrated matrices unavailable).
- 2) The input images are from **unordered** set where the sequence order are unavailable (large baselines).

Note that these two settings are common in the real applications [22]. Thus, the compared methods should able to align two uncalibrated and unordered images.

The compared methods can be divided into five categories:

- 1) **GS-based homography method.** The proposed methods **RSH** were compared to GS homography methods such as **GSH** (AutoStitch) [22].
- 2) **Commercial softwares.** We also compared our methods to well-known commercial image editing applications such as **ICE** [23] and **Photoshop** [24].
- 3) **RS video methods.** Video stabilization techniques using homography transformation by assuming scene is approximately in one plane or at infinity, are similar to the working assumption of

1. <http://www.robots.ox.ac.uk/~vgg/hzbook/code/>

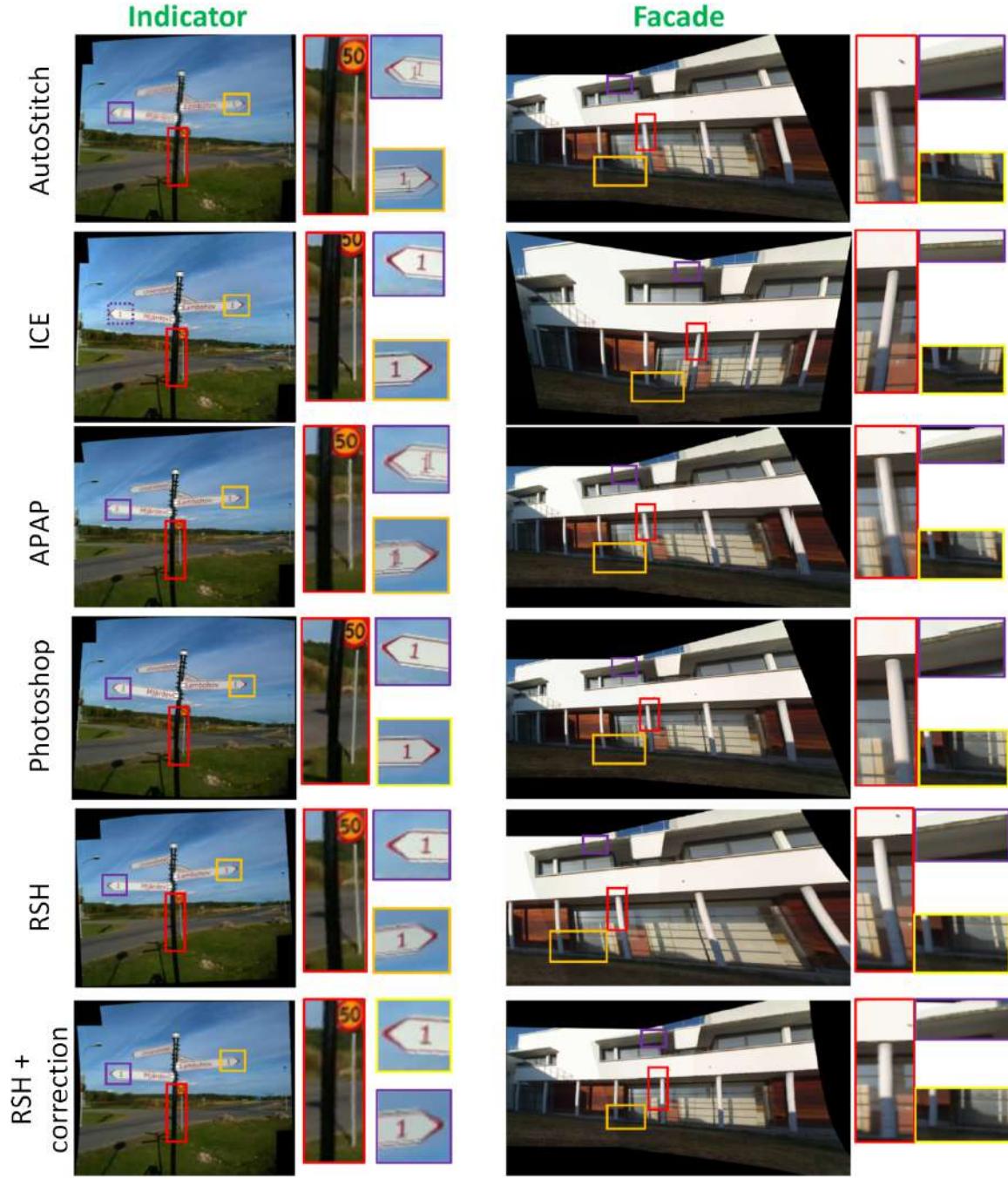


Fig. 13: Results of real uncalibrated RS images stitching with different methods.

image stitching. Therefore, video stabilization methods [4], [21], [43] for RS cameras are potentially suitable for image stitching application. However, [4] requires prior calibration and ordered video frames as input. Although [21], [43] are calibration-free methods, but they are both video-based method which violate the unordered images setting. Similarly, method in [44] generates panoramas from RS videos by assuming that the global motion of each two consecutive frames is affine. Therefore, RS video methods are unable to handle our stitching experiments.

4) Pre-rectified stitching. One solution to solve the RS stitching problem consists in performing a GS-based image stitching of pre-rectified images using RS rectification methods are such as [5],

[6], [7], [16], [27], [45]. Unfortunately, none of these methods can satisfy both above-mentioned conditions, thus can not be compared in our image stitching experiments.

5) Spatially-variant homography warping. Recent image warping techniques apply spatially-variant homography model to handle the parallax in stitching. As shown in Eq. (13), since RS homography can be explained as multiple GS homographies, image stitching methods such as **APAP** [25] and **AANAP** [26] using piecewise homographies are supposed to be able to handle the RS stitching case.

Through the discussions above, we compared the proposed method

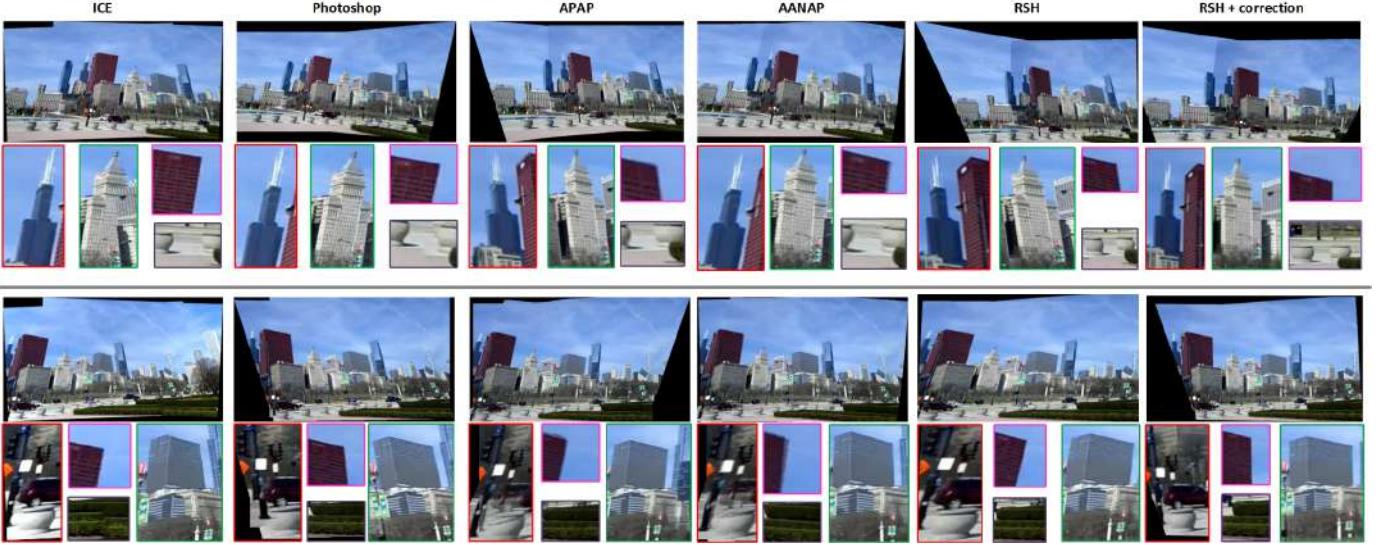


Fig. 14: Two examples of stitching real RS images from [42].

RSH to GSH, ICE, Photoshop [24] and **APAP** in RS stitching experiments.

8.2.2 Synthetic RS Images

(1) Quantitative comparison. We first compared the performances of **GSH**, **APAP**, **AANAP** and **RSH** on two synthetic RS image sequences with pure rotation from [4]: 'rot1'(camera aim changing) and 'rot2'(both changing camera aim, and in-plane rotation). In order to evaluate the stability of all the methods with different instantaneous-motions, we chose the first frame of each sequence as a reference, then we transformed and stitched all the other frames to it. We kept the number of input feature matches the same for all the three methods and calculated the average transformation errors. The results in Fig. 11 show that in the sequence 'rot' which was taken with only aim changed, all the three methods obtain similar performances while **RSH** is slightly better than **GSH**, **APAP** and **AANAP** in all the groups. However, in the sequencer 'rot2' which was taken with in-plane rotation, the transformation errors of **GSH**, **APAP** and **AANAP** increase dramatically, while **RSH** performs obviously better.

(2) Visual comparison. We visually compared the stitching results of **Autostitch**, **APAP**, **AANAP**, **ICE**, **Photoshop** and **RSH** on 'rot2' sequence by randomly choosing two frames. Fig. 12 shows that for two images with similar instantaneous-motions (first row), all methods provide visually good results. Nevertheless, for the pairs with distinct instantaneous-motions (second and third row in Fig. 12), obvious stitching defects appear (e.g. ghosting and misalignment) except with the proposed method **RSH**. Although strong RS effect remain after stitching, our method can further remove the distortion.

8.2.3 Real RS Images

(1) Real RS image sets. The first input image pair is from a RS image sequence 'indicator' [4] taken by an iPhone4. The second input is from a self-capture dataset 'facade' with strong RS effects. In Fig. 13, we can observe that **AutoStitch** produces blur on stitched images while the results from **APAP** are slightly better. The result of **ICE** in 'indicator' dataset is visually acceptable

although significant misalignments can be observed along the pole. For the 'facade' dataset, **ICE** gives a dramatically mismatched result. **Photoshop** performances are visually good in both datasets, however, some wrong alignments are still present on the overlap region such as the pole in the 'indicator' dataset and the eave in the 'facade' dataset. In contrast, **RSH** achieves the best results.

The third image pair is from [42] which captures a urban scene under fast rotation. The results in Fig. 14 show that **AutoStitch** fails in image alignment. **APAP** and **AANAP** provide blur in regions with lack of point-matches. **ICE** and **Photoshop** provide visually pleasant results, however, geometry inconsistencies or 'object's fracture' are present along the stitching seams. In contrast, the proposed method **RSH** obtains the best results.

As shown in Fig. 13 and 14, after stitching, our method can remove the distortions and offers a much more visually pleasant stitching images as final outputs.

(2) Effect of number and distribution of point matches on stitching quality. One of the main disadvantage of multiple homographies methods (**APAP** and **AANAP**) is a sufficient number of point-matches that are uniformly distributed on image plan.

In this experiment, we first evaluated the stitching quality with varying number of point-matches by decreasing the number of point correspondences. We conducted this evaluations on 'facade' dataset. Results in Fig. 15 show that the multiple local homographies (spatially-varying warping) methods such as **APAP** and **AANAP** are sensitive to the number of input point-matches. With the decreasing of input matches, the quality of stitching results with **APAP** and **AANAP** declines dramatically. In contrast, the global methods **GSH** and **RSH** show a relative high stability.

Besides, we evaluated all the stitching methods by using two real RS images from [16], which have a large overlap region. As shown in Fig. 16, the inliers between the two images distribute densely in the right part of the two images while being limited to the white facade on the left part. The stitching results show that **APAP** and **AANAP** suffer from this unbalanced point-matches distribution and provide significant distortions on the stitched regions of the 'white facade'. Slight geometrical inconsistencies are present along the stitching seams of **ICE** and **Photoshop**'s

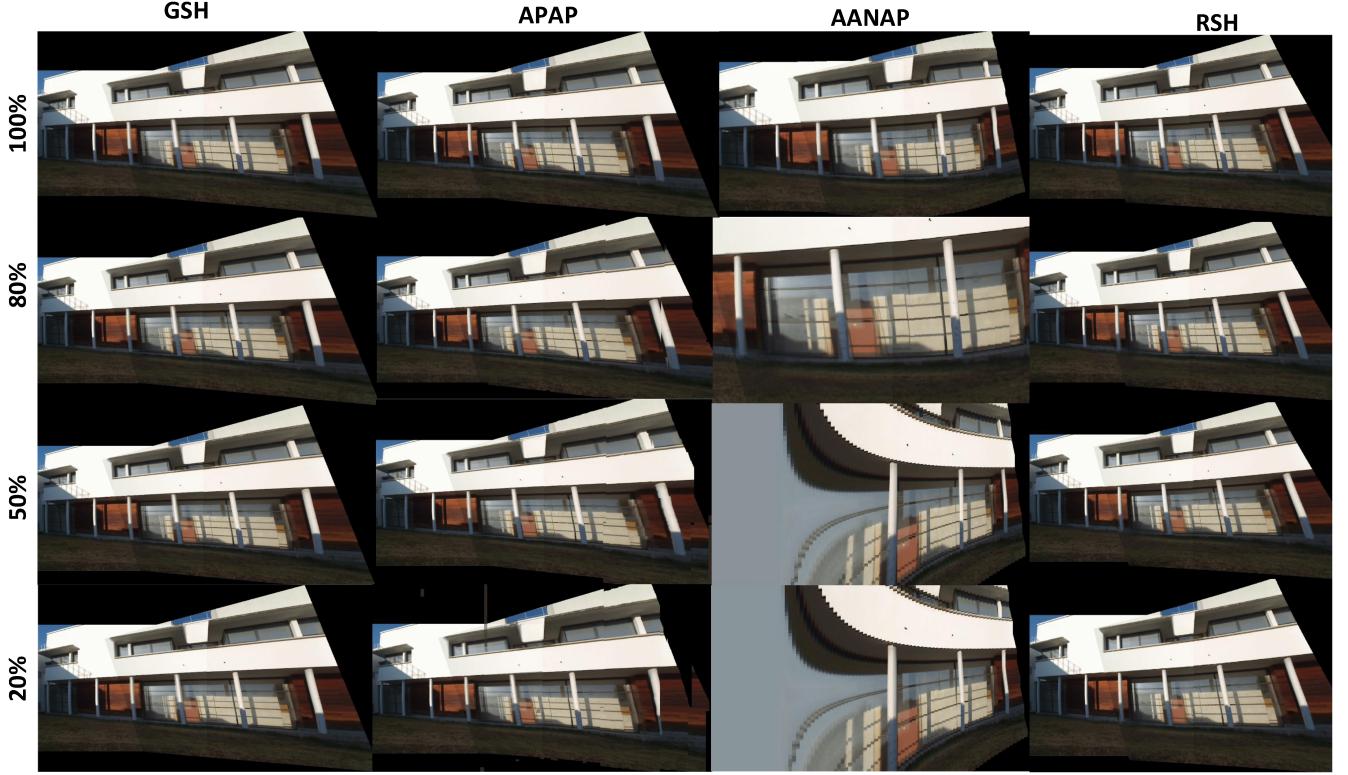


Fig. 15: Evaluation of stitching qualities by decreasing the number of input point-matches from 100% inlier to 20%.

results. In contrast, the proposed method **RSH** obtains the best result.

8.3 Running Time

The experiments were conducted on an i5 CPU at 2.8GHz with 4G RAM. On average, it took around 3.35s for **GSH**, 13.5s for **APAP** and 6.5s for **RSH**(0.05s for 13.5pt solver running per time, 0.16s for the non-linear refinement, 5.9s for the image warping, blending and RS correction). Since the proposed method was implemented in MATLAB, a significant improvement can be expected when using C++.

9 CONCLUSION

The present work is the first to address the homography for the RS case. We first defined a theoretical RS Homography matrix and proposed a 36pt solver to retrieve it from an image pair. Then we derived a practical homography matrix and the associated 13.5pt linear solver which is more suited for RANSAC based applications. The RS homography was used in two well-known homography-based applications: relative pose estimation and image stitching. The experiment results show that the proposed method is superior to the state-of-the-art techniques and some well-known commercial image editing applications.

ACKNOWLEDGMENTS

This work has been sponsored by the French government research program "Investissements d'Avenir" through the IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25), the IMobS3 Laboratory

of Excellence (ANR-10-LABX-16-01) and the RobotEx Equipment of Excellence (ANR-10-EQPX-44). This research was also financed by the European Union through the Regional Competitiveness and Employment program -2014-2020- (ERDF – AURA region) and by the AURA region.

REFERENCES

- [1] J. Linkemann and B. Weber, "Global shutter, rolling shutter—functionality and characteristics of two exposure methods (shutter variants)," *White Paper*, 2014.
- [2] C. Albl, Z. Kukelova, V. Larsson, and T. Pajdla, "Rolling shutter camera absolute pose," *PAMI*, 2019.
- [3] Y. Dai, H. Li, and L. Kneip, "Rolling shutter camera relative pose: generalized epipolar geometry," in *CVPR*, 2016.
- [4] P.-E. Forssén and E. Ringaby, "Rectifying rolling shutter video from hand-held devices," in *CVPR*, 2010.
- [5] V. Rengarajan, A. N. Rajagopalan, and R. Aravind, "From bows to arrows: Rolling shutter rectification of urban scenes," in *CVPR*, 2016.
- [6] P. Purkait, C. Zach, and A. Leonardis, "Rolling shutter correction in manhattan world," in *ICCV*, 2017.
- [7] Y. Lao and O. Ait-Aider, "A robust method for strong rolling shutter effects correction using lines with automatic feature selection," in *CVPR*, 2018.
- [8] O. Saurer, M. Pollefeys, and G. H. Lee, "A minimal solution to the rolling shutter pose estimation problem," in *IROS*, 2015.
- [9] O. Ait-Aider, N. Andreff, J. M. Lavest, and P. Martinet, "Simultaneous object pose and velocity computation using a single view from a rolling shutter camera," in *ECCV*, 2006.
- [10] Y. Lao, O. Ait-Aider, and A. Bartoli, "Rolling shutter pose and ego-motion estimation using shape-from-template," in *ECCV*, 2018.
- [11] O. Saurer, K. Koser, J.-Y. Bouguet, and M. Pollefeys, "Rolling Shutter Stereo," *ICCV*, 2013.
- [12] O. Saurer, M. Pollefeys, and G. Hee Lee, "Sparse to dense 3d reconstruction from rolling shutter images," in *CVPR*, 2016.
- [13] O. Ait-Aider and F. Berry, "Structure and kinematics triangulation with a rolling shutter stereo rig," in *ICCV*, 2009.

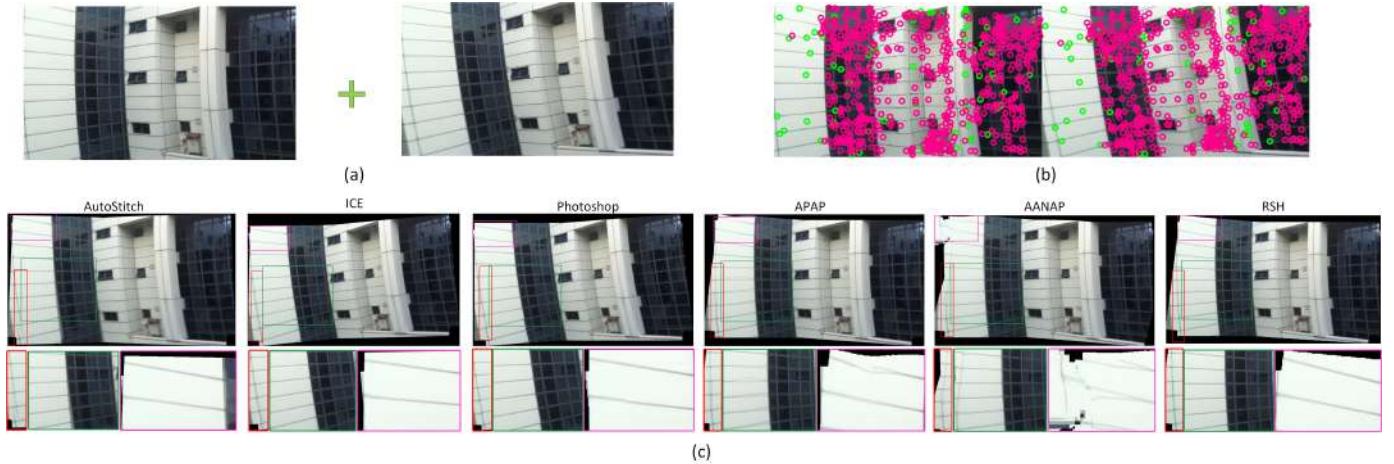


Fig. 16: (a) Two real RS images from [16] as input. (b) For fair comparisons, **APAP**, **AANAP** and **RSH** all using the same point-matches produced by GS homography RANSAC (inliers in pink, outliers in green). (c) Stitching results.

- [14] J. Hedborg, P.-E. Forssen, M. Felsberg, and E. Ringaby, "Rolling shutter bundle adjustment," in *CVPR*, 2012.
- [15] C. Albl, A. Sugimoto, and T. Pajdla, "Degeneracies in rolling shutter sfm," in *ECCV*, 2016.
- [16] B. Zhuang, L.-F. Cheong, and G. H. Lee, "Rolling-shutter-aware differential sfm and image rectification," in *ICCV*, 2017.
- [17] Y. Lao, O. Ait-Aider, and H. Araujo, "Robustified structure from motion with rolling-shutter camera using straightness constraint," *Pattern Recognition Letters*, 2018.
- [18] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *PAMI*, 2003.
- [19] D. Nistér, "An efficient solution to the five-point relative pose problem," *PAMI*, 2004.
- [20] M. Meingast, C. Geyer, and S. Sastry, "Geometric models of rolling-shutter cameras," in *OmniVis WS*, 2005.
- [21] M. Grundmann, V. Kwatra, D. Castro, and I. Essa, "Calibration-free rolling shutter removal," in *ICCP*, 2012.
- [22] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *IJCV*, 2007.
- [23] "Image composite editor - microsoft research." [Online]. Available: <https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/>
- [24] "Adobe photoshop cc." [Online]. Available: <https://www.adobe.com/products/photoshop.html>
- [25] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving dlt," in *PAMI*, 2014.
- [26] C.-C. Lin, S. U. Pankanti, K. Natesan Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *CVPR*, 2015.
- [27] S. Vasu, M. M. MR, and A. Rajagopalan, "Occlusion-aware rolling shutter rectification of 3d scenes," in *CVPR*, 2018.
- [28] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [29] Z. Zhou, H. Jin, and Y. Ma, "Robust plane-based structure from motion," in *CVPR*, 2012.
- [30] O. Saurer, P. Vasseur, R. Boutteau, C. Demonceaux, M. Pollefeys, and F. Fraundorfer, "Homography based egomotion estimation with a common direction," *PAMI*, 2017.
- [31] B. Klingner, D. Martin, and J. Roseborough, "Street view motion-from-structure-from-motion," in *ICCV*, 2013.
- [32] E. Ito and T. Okatani, "Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion," in *CVPR*.
- [33] P. Purkait and C. Zach, "Minimal solvers for monocular rolling shutter compensation under ackermann motion," in *WACV*, 2018.
- [34] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, 1981.
- [35] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-d vision: from images to geometric models*. Springer, 2012.
- [36] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control," Ph.D. dissertation, INRIA, 2007.
- [37] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics (TOG)*, vol. 2, no. 4, pp. 217–236, 1983.
- [38] M. Brown, D. G. Lowe *et al.*, "Recognising panoramas," in *ICCV*, vol. 3, 2003, p. 1218.
- [39] H. Stewénius, "Gröbner basis methods for minimal problems in computer vision," Ph.D. dissertation, Lund University, 2005.
- [40] Z. Kukelova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," in *ECCV*, 2008.
- [41] V. Larsson, K. Astrom, and M. Oskarsson, "Efficient solvers for minimal problems by syzygy-based reduction," in *CVPR*, 2017.
- [42] C. Jia and B. L. Evans, "Probabilistic 3-d motion estimation for rolling shutter video rectification from visual and inertial measurements," in *MMSP*, 2012, pp. 203–208.
- [43] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled camera paths for video stabilization," *ACM Transactions on Graphics (TOG)*, 2013.
- [44] W.-H. Cho and T.-C. Kim, "Cis video panoramic image," in *2012 IEEE 16th International Symposium on Consumer Electronics*. IEEE, 2012, pp. 1–4.
- [45] V. Rengarajan, Y. Balaji, and A. Rajagopalan, "Unrolling the shutter: Cnn to correct motion distortions," in *CVPR*, 2017.



Yizhen Lao Yizhen Lao obtained Master's degree in Geo-information Science and Earth Observation from University of Twente and Bachelor's degree in Spatial-Informatics and Digitalized Technology from Wuhan University. Currently, he is pursing his PhD degree in computer vision at Institut Pascal, Université Clermont Auvergne / CNRS. He focuses on 3D vision with non-global acquisition cameras (e.g. Rolling Shutter cameras).



Omar Ait-Aider Omar Ait-Aider is an Associate Professor and the Head of the Bachelor degree in Mechatronics at University of Clermont-Auvergne. He received his Master's Degree on "Autonomous Robotics" at Pierre et Marie Curie University in Paris, then his PhD on "Computer Vision for Robotics" at the University of Evry Val d'Essonne in France. Since September 2006, he is member of the "Image, Perception Systems and Robotics" group within the Institut Pascal-CNRS. His research is focused on geometrical camera modelling and 3D vision. He received the Honorable Mention Best Paper Award at the European Conference on Computer Vision ECCV'2006.