Rolling Shutter Pose and Ego-motion Estimation using Shape-from-Template

Yizhen Lao, Omar Ait-Aider and Adrien Bartoli

Institut Pascal, Université Clermont Auvergne / CNRS, France
lyz91822@gmail.com, omar.ait-aider@uca.fr, adrien.bartoli@gmail.com

Abstract. We propose a new method for the absolute camera pose problem (PnP) which handles Rolling Shutter (RS) effects. Unlike all existing methods which perform 3D-2D registration after augmenting the Global Shutter (GS) projection model with the velocity parameters under various kinematic models, we propose to use local differential constraints. These are established by drawing an analogy with Shape-from-Template (SfT). The main idea consists in considering that RS distortions due to camera ego-motion during image acquisition can be interpreted as virtual deformations of a template captured by a GS camera. Once the virtual deformations have been recovered using SfT, the camera pose and ego-motion are computed by registering the deformed scene on the original template. This 3D-3D registration involves a 3D cost function based on the Euclidean point distance, more physically meaningful than the re-projection error or the algebraic distance based cost functions used in previous work. Results on both synthetic and real data show that the proposed method outperforms existing RS pose estimation techniques in terms of accuracy and stability of performance in various configurations.

Keywords: Rolling Shutter; Pose estimation; Shape-from-Template

1 Introduction

Many modern CMOS cameras are equipped with Rolling Shutter (RS) sensors which are relatively low-cost and electronically advantageous compared to Global Shutter (GS) ones. However, in RS acquisition mode, the pixel rows are exposed sequentially from the top to the bottom of the image. Therefore, images captured by moving RS cameras produce distortions (e.g. wobble, skew), which defeat the classical GS geometric models in 3D computer vision. Thus, new methods adapted to RS cameras are strongly desired. Recently, many methods have been designed to fit RS camera applications such as object pose calculation [1–3], 3D reconstruction from stereo rigs [3–5], bundle adjustment [6, 7], relative pose estimation [8], dense matching [9, 4] and degeneracy understanding [10, 11].

Camera pose estimation (PnP) is the problem of calculating the pose of a calibrated camera from n 3D-2D correspondences. Camera pose estimation is important and extensively used in Simultaneous Localization And Mapping (SLAM) for robotics, object or camera localization and Augmented Reality (AR). Most existing works focus on solving the minimal problem based on the GS model [12–15] with at least three point matches. Given such a minimal solution, RANSAC and non-linear optimization are two frameworks to further improve robustness and accuracy [16]. However, estimating the RS camera pose with the GS model does not give satisfactory results [17].

A few works focus on RS camera pose estimation [18, 17, 19]. These all try to extend GS-based PnP solutions by incorporating camera ego-motion in the projection model. In contrast, we provide a completely new perspective in RS projection and propose a novel solution to estimate RS camera pose and ego-motion simultaneously (**RS-PEnP**).

Our solution is based on an analogy with Shape-from-Template (SfT). This is the problem of reconstructing the shape of a deformable surface from a 3D template and a single image [20]. We show that theoretically RS image distortions caused by camera ego-motion can be expressed as virtual deformations of 3D shapes captured by a GS camera. Thus, the idea is to first retrieve virtual shape deformations using SfT, and then to re-interpret these deformations as RS effects by estimating camera ego-motion thanks to a new 3D-3D registration technique. By transforming the RS PnP problem into a 3D-3D registration problem, we show that our RS-PEnP solution is more robust and stable than existing works [17] because the constraints to be minimized are more physically meaningful and are all expressed in the same metric dimension.

1.1 Related Work and Motivations

One of the key issues in solving RS geometric problems is incorporating feasible camera ego-motion into projection models. Saurer et al. [18] propose a minimal solver to estimate RS camera pose based on the translation-only model with at least 5 3D-2D correspondences. However, this solution is limited to specific scenarios, such as a forward moving vehicle. It is not feasible for the majority of applications such as a hand-held camera, a drone or a moving robot, where ego-rotation contributes significantly to RS effects [7, 21]. Albl et al. [19] propose another minimal solver, which requires at least 5 3D-2D matches too. It is based on a uniform ego-motion model. Nevertheless, it also requires the assistance of inertial measurement units (IMUs), which makes the algorithm dependent on additional sensors. Albl et al. also propose a minimal and non-iterative solution to the RS-PEnP problem called R6P [17], which can achieve higher accuracy than the standard P3P [12] by using an approximate doubly-linearized model. The approximation requires that the rotation between camera and world frames is small. Therefore, all 3D points need to be rotated first to satisfy the double-linearization assumption based on a rough estimation from IMU measurements or P3P. This pre-processing step makes R6P suffer from dependencies on additional sensors or the risk that P3P gives a non satisfactory rough estimate. Besides, R6P gives up to 20 feasible solutions and no flawless recipe is provided to choose the right one, which may lead to unstable performances.

Magerand et al. [22] present a polynomial projection model for RS cameras and propose the constrained global optimization of its parameters by means

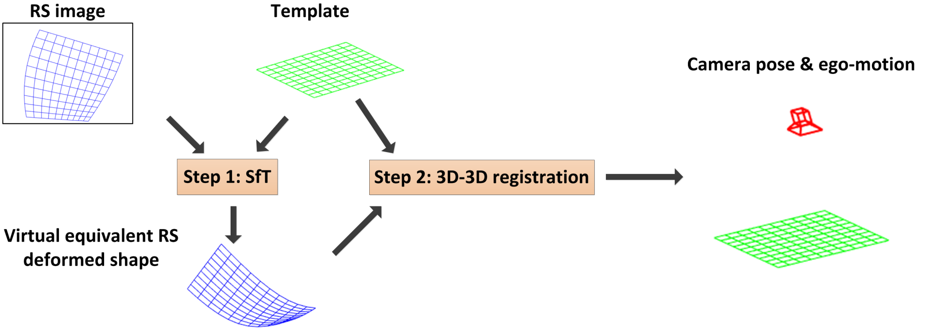
RS image
Template
Camera pose & ego-motion
Step 1: SfT
Step 2: 3D-3D registration
Virtual equivalent RS deformed shape

Fig. 1. An overview of the proposed pose and ego-motion estimation method: *Step 1:* Given an RS image and a known 3D template, we reconstruct the virtually deformed shape using SfT. *Step 2:* By performing 3D-3D registration between the virtually deformed shape and the template, RS camera pose and ego-motion are obtained simultaneously.

of a semidefinite programming problem obtained from the generalized problem of moments method. Contrarily to other methods, their optimization does not require an initialization and can be considered for automatic feature matching in a RANSAC framework. Unfortunately, the resolution is left to an automatic but computationally expensive solver.

In summary, a new efficient and stable solution to estimate the pose and ego-motion of an RS camera under general motion, without the need for other sensors, is still absent from the literature. Such a solution is highly required by many potential applications.

1.2 Contribution and Paper Organization

The main contributions of this paper are:

- We show and prove for the first time that RS effects can be explained by the GS-based projection of a virtually deformed shape. Thus, we show the analogy between the SfT and RS-PEnP problems.
- We propose a novel RS-PEnP method, illustrated in Fig. 1, which first recovers the virtual template deformation using SfT and then computes the pose and velocity parameters using 3D-3D registration.

We first introduce the RS projection model and the formulation of the RS camera pose problem in section 2. Then we give a brief introduction to the SfT problem in section 3. The links between the SfT and RS-PEnP problems are analyzed in section 4. In section 5, we show how to retrieve RS camera pose and ego-motion by using 3D-3D registration. The evaluation of the proposed method and conclusions are presented in sections 6 and 7.

2 The RS Pose and Ego-motion Problem

2.1 The RS Projection Model

In the static case, an RS camera is equivalent to a GS one. It follows a classical pinhole camera projection model defined by the intrinsic parameters matrix $\mathbf{K}$, rotation $\mathbf{R} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$ between world and camera coordinate systems [23]:

$$\mathbf{q}_i = \mathit{\Pi}^{GS}(\mathbf{K}[\mathbf{R} \quad \mathbf{t}]\widetilde{\mathbf{P}}_i) = \mathit{\Pi}^{GS}(\mathbf{K}\mathbf{Q}_i) \tag{1}$$

where $\mathit{\Pi}^{GS}$ is the GS projection operator defined as $\mathit{\Pi}^{GS}([X\ Y\ Z]^\top) = \frac{1}{Z}[X\ Y]^\top$, $\widetilde{\mathbf{P}}_i$ are the homogeneous coordinates of a 3D point $\mathbf{P}_i = [X_i, Y_i, Z_i]^\top$ in world coordinates, transformed by camera pose into camera coordinates as $\mathbf{Q}_i$. Finally, $\mathbf{q}_i = [u_i, v_i]^\top$ is its projection in the image.

For a moving RS camera, during frame exposure, each row will be captured in turn and thus with a different pose, yielding a new projection operator $\mathit{\Pi}^{RS}$. Thus, Eq. (1) becomes:

$$\mathbf{q}_i = \mathit{\Pi}^{RS}(\mathbf{K}\mathbf{Q}_i) = \mathit{\Pi}^{GS}(\mathbf{K}\mathbf{Q}_i^{RS}) = \mathit{\Pi}^{GS}(\mathbf{K}[\mathbf{R}(v_i) \quad \mathbf{t}(v_i)]\widetilde{\mathbf{P}}_i) \tag{2}$$

where $\mathbf{R}(v_i)$ and $\mathbf{t}(v_i)$ define the camera pose when the image row of index v_i is acquired. Therefore, a static 3D point $\mathbf{P}_i$ in world coordinates is transformed into $\mathbf{Q}_i^{RS}$, instead of $\mathbf{Q}_i$, in camera coordinates.

2.2 RS Pose and Ego-motion Estimation (RS-PEnP)

Except [24], all existing methods for RS are based on augmenting the projection model by the rotational and translational velocity parameters during image acquisition. Considering that the scanning time for one frame is generally very short, different kinematic models are considered in order to express $\mathbf{R}(v_i)$ and $\mathbf{t}(v_i)$. Unfortunately, these additional parameters bring non-linearity in the projection model. A compromise should then be found between the accuracy of the kinematic model and the possibility to find an elegant and efficient solution for the RS-PEnP problem. A realistic simplified model is the uniform motion during each image acquisition (i.e. constant translational and rotational speed).

3 Shape-from-Template

SfT refers to the task of template-based monocular 3D reconstruction, which estimates the 3D shape of a deformable surface by using different physic-based deformation rules [25, 20]. Fig. 2 illustrates the geometric modeling of SfT. A 3D template $\tau \subset \mathbb{R}^3$ transforms to the deformed shape $S \subset \mathbb{R}^3$ by a 3D deformation $\mathit{\Psi} \in C^1(\tau, \mathbb{R}^3)$. If $\mathit{\Omega} \subset \mathbb{R}^2$ is a 2D space obtained by flattening a 3D template τ, thus, an unknown deformed embedding $\varphi \subset C^1(\mathit{\Omega}, \mathbb{R}^3)$ maps a 2D point $\mathbf{p} \in \mathit{\Omega}$ to $\mathbf{Q} \in S$. Finally, $\mathbf{Q}$ can be projected onto an image point $\mathbf{q} \in I$ by a known

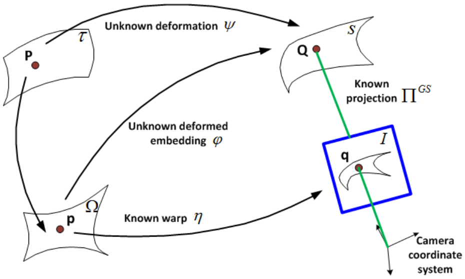
Unknown deformation Ψ
$\mathcal{T}$
P
Q
S
Known projection Π^{GS}
Unknown deformed embedding φ
I
q
Ω
p
Known warp η
Camera coordinate system

Fig. 2. Geometric modeling of SfT.

GS-based projection function Π^{GS}. The known transformation between Ω and I is denoted as η. It is obtained from 3D-2D point correspondences using Bsplines as in [26]. The goal of SfT is to obtain the deformed surface S given $\mathbf{p}$, $\mathbf{q}$ and the first order derivatives of the optical flow at $\mathbf{p}$, namely $\frac{\partial \eta}{\partial p}(\mathbf{p})$.

The deformation constraints used to solve SfT can be categorized as:

- Isometric deformation. The geodesic distances are preserved by the deformation [20, 25–28]. This assumption commonly holds for paper, cloth and volumetric objects.
- Conformal deformation. The isometric constraint can be relaxed to conformal deformation, which preserves angles and possibly handles isotropic extensible surfaces such as a balloon [20].
- Elastic deformation. Linear [29, 30] or non-linear [31] elastic deformations are used to constrain extensible surfaces. Elastic SfT does not have local solution, in contrast to isometric SfT, and requires boundary condition to be available, as a set of known 3D surface points.

4 Moving Object under RS or Deformed Template under GS?

4.1 An Equivalence between RS Projection and Surface Deformation

The main idea in this paper is that distortions in RS images caused by camera ego-motion can be expressed as the virtual deformation of a 3D shape captured by a GS camera. We first model the GS projection of a known 3D shape after a deformation Ψ:

$$\mathbf{m}_i = \Pi^{GS}(\mathbf{K}\Psi(\mathbf{P}_i)) \tag{3}$$

If we define the deformation as $\Psi^{RS}(\mathbf{P}_i) = \mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i)$, Eq. (3) becomes similar to Eq. (2):

$$\mathbf{m}_i = \Pi^{GS}(\mathbf{K}\Psi^{RS}(\mathbf{P}_i)) = \Pi^{GS}(\mathbf{K}(\mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i))) = \Pi^{RS}(\mathbf{K}\mathbf{Q}_i) \tag{4}$$

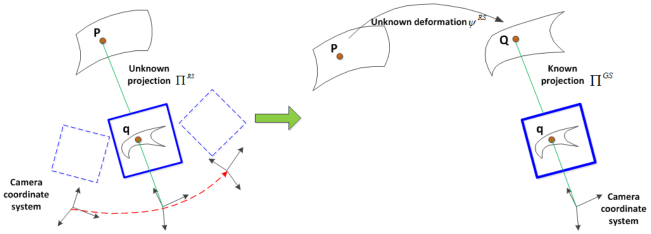

Fig. 3. Equivalence between the RS projection of a rigid object and a GS projection of a virtually deformed template.

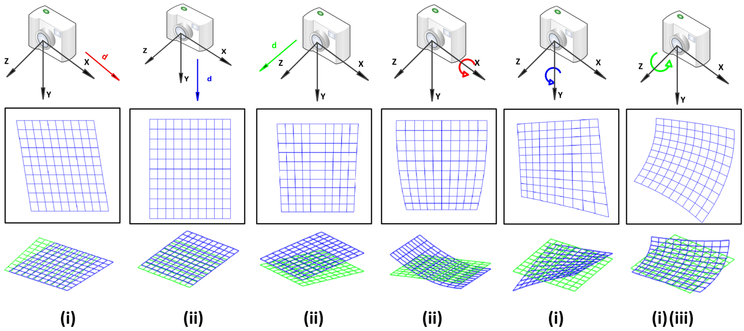

Fig. 4. The 3D template shapes (green) captured by a RS camera under different atomic ego-motions (first row) yield distorted RS images (second row). The exact same images are also obtained as the projection of the corresponding virtually deformed 3D shapes (blue) into a GS camera (third row). The type of corresponding virtual deformations are also given, see main text for details.

Eq. (4) and Fig. 3 show that 3D shapes observed by a moving an RS camera are equivalent to corresponding deformed 3D shapes filmed by a GS camera. We name this virtual corresponding deformation Ψ^{RS} as the ***equivalent RS deformation*** and the virtually deformed shape $\Psi^{RS}(\mathbf{P}_i)$ as the ***equivalent RS deformed shape***.

4.2 Reconstruction of the Virtual RS Deformed Shape

After showing the link between the RS-PEnP and SfT problems, we focus on how to reconstruct the equivalent RS deformed shape by using SfT. Since the assumption on the physical properties of the template plays a crucial role in solving the SfT problem we should determine which one of the deformation constraints can best describe the equivalent RS deformation.

Any RS ego-motion can be regarded as a combination of six atomic ego-motions: translation along the X ($\mathbf{d}_x$), Y ($\mathbf{d}_y$), Z ($\mathbf{d}_z$) axes and rotation about the X ($\boldsymbol{\omega}_x$), Y ($\boldsymbol{\omega}_y$), Z ($\boldsymbol{\omega}_z$) axes. Fig. 4 shows RS images and equivalent RS deformed shapes produced by different types of RS ego-motions. Albl et al. [19] and Rengarajan et al. [32] illustrated four different types of RS effects (2D deformations) produced by camera ego-motion. Differently, we focus on virtual 3D deformations instead. Fig. 4 also shows that the corresponding virtual deformations caused by different camera ego-motions can be summarized into three types, by assuming a vertical scanning direction of the 3D template:

(i) Horizontal wobble: Translation along the x-axis, rotation along the y-axis and z-axis create surface wobble along the horizontal direction (perpendicular to the scan direction). In such cases, the distances are preserved only along the horizontal direction while the angles change during the deformation.

(ii) Vertical shrinking/extension: Translation along the y-axis or rotation along the x-axis produce a similar effect, which shrinks or extends the 3D shape along the scan direction (vertical). This deformation preserves the distances along the horizontal direction but changes the angles. Thus, unlike an elastic deformation, stretching the surface in the vertical direction will not introduce a compression in the horizontal direction.

(iii) Vertical wobble: Beside horizontal wobble, rotation along the z-axis also leads to wobble in the vertical direction. The distances along the horizontal direction remain unchanged while the angles vary dynamically.

It is important to notice that the virtual deformations do not follow any classical physics-based SfT surface models such as isometry, conformity or elasticity: isometric surface deformation preserves the distances along all directions while the equivalent RS distortion only preserves the distances along the horizontal direction. The conformal deformation is a relaxation of the isometric model, which allows local isotropic scaling and preserves the angles during deformation. However, it cannot describe how the virtual deformation angles change. The elastic surface stretches in one direction and generally produces extension in the orthogonal direction. In contrast, no shrinking or extension occurs along the horizontal direction during the equivalent RS deformation.

We focus on reconstructing the equivalent RS deformed shape based on the isometric and conformal deformations for the following reasons:

- The isometric constraint holds along the horizontal direction on the 3D shapes. Since the image acquisition time is commonly short, the effects of extension and compression of the 3D shape are limited, which makes the isometric model work in practice. Alternatively, the conformal model can reconstruct extensible 3D shapes [20]. Thus, the conformal model as a relaxation of the isometric model can be theoretically considered a better approximation to the equivalent RS deformation.

- A complex equivalent RS deformed shape will be produced if an RS camera is under general ego-motion, which is the composition of six types of atomic ego-motions. Therefore, different surface patches on the shape could be under different 3D deformations. Importantly, the isometric and conformal SfT solutions we used from [20] exploit **local** differential constraints and recover the local deformation around each point on the shape independently. The assumption we

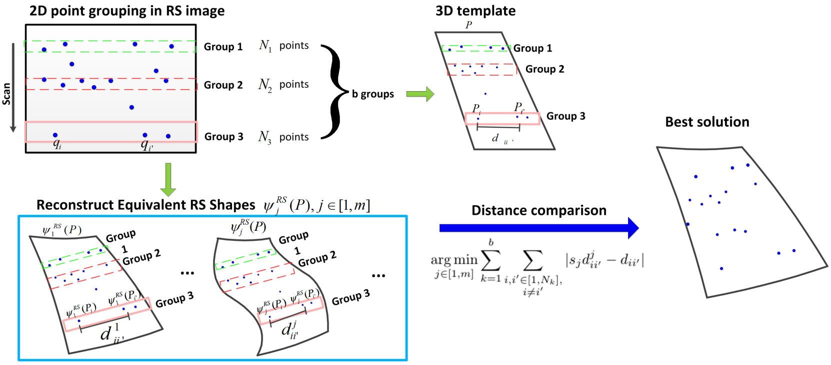

Fig. 5. Choosing the best equivalent RS shape from conformal SfT.

implicitly make is thus that the camera projection is GS in each neighbourhood. This turns out to be a very mild and valid assumption in practice.

- Analytical solutions to SfT using the isometric and conformal models are reported in [20], which are therefore faster and show the potential to form real-time applications [27]. In contrast, the existing solutions to the elastic model are made slower [29, 30] and require boundary conditions unavailable in RS-PEnP.

Isometric deformation. Bartoli et al. showed that only one solution exists to isometric surface reconstitution from a single view and proposed the first analytical algorithm [20]. A stable solution framework for isometric SfT has been proposed later [28]. Thanks to the existing isometric algorithms, we can stably and efficiently obtain a single reconstruction of equivalent RS deformed shape $\Psi^{RS}(\mathbf{P}_i)$.

Conformal deformation. Contrarily to the isometric case, conformal-based SfT theoretically yields a small, discrete set of solutions (at least two) and a global scale ambiguity [20]. Thus, we obtain multiple reconstructed equivalent RS deformed shapes by using the analytical SfT method under the conformal constraint. However, only one reconstruction is close to the real equivalent RS deformed shape $\Psi^{RS}(\mathbf{P}_i)$. Therefore, we pick up the most practically reasonable reconstruction based on distance preservation along the horizontal direction. We assume that a total of m reconstructed shapes $\left\{\Psi_j^{RS}(\mathbf{P}), \quad j = 1, 2..., m\right\}$ are obtained. As shown in Fig. 5 the 2D points located close to each other in the scanning direction in the image are segmented into b groups $\mathbb{G}_k, k \in [1, b]$ of N_k points. In the experiments, we group two 2D points into the same group if their difference of row index is lower than a threshold (experimentally set as 5 pixels). Then, we calculate a global scale factor s_j of each reconstructed equivalent RS deformed shape to the template by using $s_j = \frac{2}{n(n-1)} \sum_{i,i' \in [1,n], i \neq i'} d_{ii'}/d_{ii'}^j$, where $d_{ii'}$ is the euclidean distance between 3D points $\mathbf{P}_i$ and $\mathbf{P}_{i'}$ and $d_{ii'}^j$ is the euclidean distance of the corresponding reconstructed 3D points $\Psi_j^{RS}(\mathbf{P}_i)$ and $\Psi_j^{RS}(\mathbf{P}_{i'})$. We choose $i, i' \in [1, n]$ randomly and calculate the average value.

Finally, we choose the reconstruction $\Psi_j^{RS}(\mathbf{P})$ with the smallest sum of distance differences along the horizontal direction between each equivalent RS deformed shapes ${}^{x}d_{ii'}^{j}$ and known 3D template ${}^{x}d_{ii'}$ as the best solution:

$$\underset{j\in[1,m]}{\arg\min} \sum_{k=1}^{b} \sum_{\substack{i,i'\in[1,N_k],\\ i\neq i'}} |s_j {}^{x}d_{ii'}^{j} - {}^{x}d_{ii'}| \tag{5}$$

5 Camera Pose and Ego-motion Computation

5.1 Kinematic Model and RS Projection

Since the acquisition time of a frame is commonly short, one can generally assume a uniform kinematic model (with constant translational and rotational velocities). Moreover, by considering small rotation angles, we obtain the so-called linearized model, which has been used in many applications [22, 17, 8, 10].

By using the linearized RS ego-motion model, the rotation $\mathbf{R}(v_i)$ and translation $\mathbf{t}(v_i)$ of the i^{th} row in Eq. (2) become:

$$\begin{aligned} \mathbf{R}(v_i) &= (\mathbf{I} + [\omega]_\times v_i)\mathbf{R}_0 \\ \mathbf{t}(v_i) &= \mathbf{t}_0 + \mathbf{d}v_i \end{aligned} \tag{6}$$

where $\mathbf{R}_0$ and $\mathbf{t}_0$ are the rotation and the translation of the first row, which we set as the reference pose for the frame, $\mathbf{d}$ and $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^\top$ are the translational and rotational velocities respectively. Thus, the rotation during acquisition can be defined by Rodrigues's formula as $\mathbf{a}\mathbf{a}^\top(1-\cos(v_i\omega)) + \mathbf{I}\cos(v_i\omega) + [\mathbf{a}]_\times \sin(v_i\omega)$, where $\omega = |\boldsymbol{\omega}|$, $\mathbf{a} = \boldsymbol{\omega}/\omega$. With the assumption of short acquisition time, Rodrigues's formula can be simplified as $\mathbf{I} + v_i[\boldsymbol{\omega}]_\times$ by using the first order Taylor expansion, where $[\boldsymbol{\omega}]_\times$ is the skew-symmetric matrix of $\boldsymbol{\omega}$.

5.2 3D-3D Registration

After obtaining the equivalent RS shape $\Psi^{RS}(\mathbf{P})$, we register the virtually deformed shape to the known 3D template $\mathbf{P}$ using the RS ego-motion model. By iteratively minimizing the distance errors between the known 3D template and the reconstructed equivalent RS shape, we can obtain the camera pose and ego-motion parameters simultaneously:

$$\underset{\mathbf{R}_0,\mathbf{t}_0,\boldsymbol{\omega},\mathbf{d}}{\arg\min} \sum_{i=1}^{n} \left\| \mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i) - \Psi^{RS}(\mathbf{P}_i) \right\| \tag{7}$$

where $\mathbf{R}_0$ and $\mathbf{t}_0$ are initialized using a classical PnP method [34]. Actually, we slightly abused the term ‘registration’ to mean that the 3D points of the virtually deformed surface are fitted with the corresponding 3D points of the template. This can be seen as a registration where the recovered parameters are not a mere rigid transformation but a local motion with constant velocity.

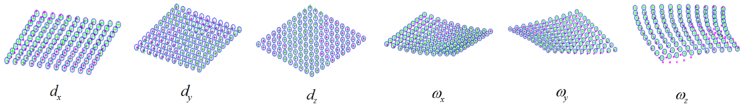

Fig. 6. Reconstructed equivalent RS deformed shapes by **AnIRS** (magenta points) and **AnCRS** (green crosses) compared to ground truth structure (blue circles) under six types of camera ego-motion.

The ego-motion parameters ($\boldsymbol{\omega}$,$\mathbf{d}$) are initialized by the following two steps: (1) Group image points into sets of vertically close points (so that the RS effect can be neglected) and run P3P for each set. (2) Initialize $\mathbf{d}$ and $\boldsymbol{\omega}$ by computing the relative translation and rotation between groups and dividing by the scan time. Alternatively, we can operate in the same procedure by grouping the points of the deformed surface into subsets of close 3D points, which are registered by 3D point transformations [35]. However, in many practical situations, it is more convenient and more efficient to set the initial values of $\mathbf{d}$ and $\boldsymbol{\omega}$ to 0, which in our experiments always allowed convergence toward the correct solution.

6 Experiments

In our experiments, the analytical isometric solution [28][1] (**AnIRS**) and analytical conformal solution (**AnCRS**) [20][1] are used to reconstruct the equivalent RS deformed shape of both synthetic and real planar and non planar templates under isometric and conformal constraints respectively. The Levenberg-Marquardt algorithm is used in the non-linear pose and ego-motion estimation from Eq. (7).

6.1 Synthetic Data

We simulated a calibrated pin-hole camera with 640 × 480 px resolution and 320 px focal length. The camera was located randomly on a sphere with a radius of 20 units and was pointing to a simulated surface (10×10 units) with varying average scanning angles from 0 to 90 deg. We randomly drew n points on the surface to form the 3D template. Random Gaussian noise with standard deviation σ was also added to the 2D projected points $\mathbf{q}$.

Recovering the equivalent RS deformed shape. We first evaluate the reconstruction accuracy of **AnIRS** and **AnCRS** on the equivalent RS deformed shape. Since the types of deformation depend on the type of RS ego-motion, we measure the mean and standard deviation of distances between the reconstructed 3D points and the corresponding points on the 3D template under six atomic ego-motion types (section 4.2). For each motion type, we run 200 trials to obtain statistics. We varied the number of 3D-2D matches from 10 to 121 and

[1] http://igt.ip.uca.fr/∼ab/Research

Table 1. Mean ($|e_I|$, $|e_C|$) and standard deviation (σ_I, σ_C) of reconstruction errors (expressed in units) of the equivalent RS deformed shape by **AnIRS** and **AnCRS** under six types of camera ego-motion.

	d_x	d_y	d_z	ω_x	ω_y	ω_z
$\lvert e_I \rvert$	0.0130283	0.0113629	**0.0001183**	0.0023273	0.0020031	0.1338190
$\lvert e_C \rvert$	**0.0040963**	**0.0052104**	0.0009037	**0.0000921**	**0.0008493**	**0.0008417**
σ_I	0.0001810	0.0000943	**0.0000014**	0.0000834	0.0007209	0.0393570
σ_C	**0.0000318**	**0.0000529**	0.0000310	**0.0000206**	**0.0003639**	**0.0001201**

used a noise level $\sigma = 1$ px. At each trial, the ego-motion speed was randomly set as follows: translational speed varying from 0 to 3 units/frame and rotational speed varying from 0 to 20 deg/frame.

The results in Fig. 6 show that both **AnIRS** and **AnCRS** provide stable and high accuracy results for the equivalent RS deformed shape reconstruction. The quantitative evaluation in Table 1 demonstrates that **AnCRS** generally performs better than **AnIRS**. Specifically, it indicates that the advantages of **AnCRS** are significant in the cases of ego-rotation along the x or z-axis. The only exception is in translation along the z-axis, where the equivalent RS deformation is with relatively smaller extension/shrinking than other types. Thus, **AnIRS** gives better results than **AnCRS**. However, all observations confirm our analysis in section 4.2 that conformal surfaces can better model the extensibility of equivalent RS deformation generally.

Pose estimation. We compared **AnIRS** and **AnCRS** in camera pose estimation with both the GS PnP solution **GS-PnP**[2] [13] and the RS-PEnP solution **RS-PnP**[3] which uses R6P [17]. Since the ground truth of camera poses are known, we measured the absolute error of rotation (deg) and translation (units).

- ***Accuracy vs ego-motion speed.*** We fixed the number of 3D-2D matches to 60 and noise level to $\sigma = 1$ px. We increased the translational speed and rotational speed from 0 to 3 units/frame and 20 deg/frame gradually. At each configuration, we run 100 trials with random velocity directions and measured the average pose errors. The results in Fig. 7(a,b) show that both **AnIRS** and **AnCRS** provide significantly more accurate estimates of camera orientation and translation with all ego-rotation configurations (ω_x, ω_y and ω_z) compared to **GS-PnP** and **RS-PnP**. Under three ego-translations, **AnIRS** and **AnCRS** show an obvious superiority in camera rotation estimation, and perform slightly better in camera translation estimation than **RS-PnP**. As expected, GS-based **GS-PnP** fails in pose estimation once the ego-motion is strong. In contrast, **RS-PnP** achieves better results in translation than **GS-PnP**, but both of them provide an inaccurate estimate for camera rotation to the same extent.
- ***Accuracy vs image noise.*** In this experiment, we evaluated the robustness of the four solutions against different noise levels. Thus, we fixed the camera with

[2] estimateWorldCameraPose function in MATLAB

[3] http://cmp.felk.cvut.cz/∼alblcene/r6p

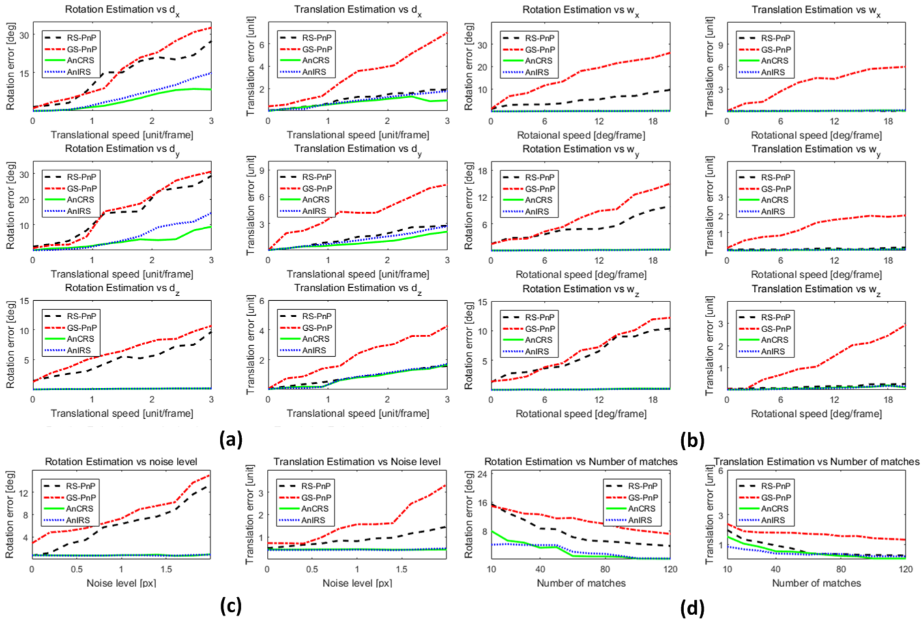

Fig. 7. Pose estimation errors for **AnIRS**, **AnCRS**, the GS **GS-PnP** and **RS-PnP** under different ego-translations (a), ego-rotations (b), image noise levels (c) and number of matches (d).

translational and rotational speed at 1 unit/frame and 5 deg/frame. Random noise with levels varying from 0 to 2 px were added to the 60 image points. The results in Fig. 7(c) show that both **AnIRS** and **AnCRS** are robust to the increasing image noise. In contrast, **GS-PnP** and **RS-PnP** are relatively sensitive to image noise.

• ***Accuracy vs number of matches.*** The number of 3D-2D matches has a great impact on the PnP problem. Therefore, we evaluated the performance of the proposed method with different numbers of 3D-2D matches. The camera was fixed with translational and rotational speed at 1 unit/frame and 5 deg/frame. The image noise level was set to 1 px. Then we increased the number of matches from 10 to 120. The results in Fig. 7(d) show that the estimation accuracy of all four methods increases with the number of matches. However **AnIRS** and **AnCRS** provides better results in both rotation and translation estimation compared to **GS-PnP** and **RS-PnP**.

6.2 Real Data

Augmented Reality with an RS video. The four methods have been further evaluated by using real RS images. A planar marker providing 64 3D-2D matches was captured by a hand-held logitech webcam. Strong RS effects are

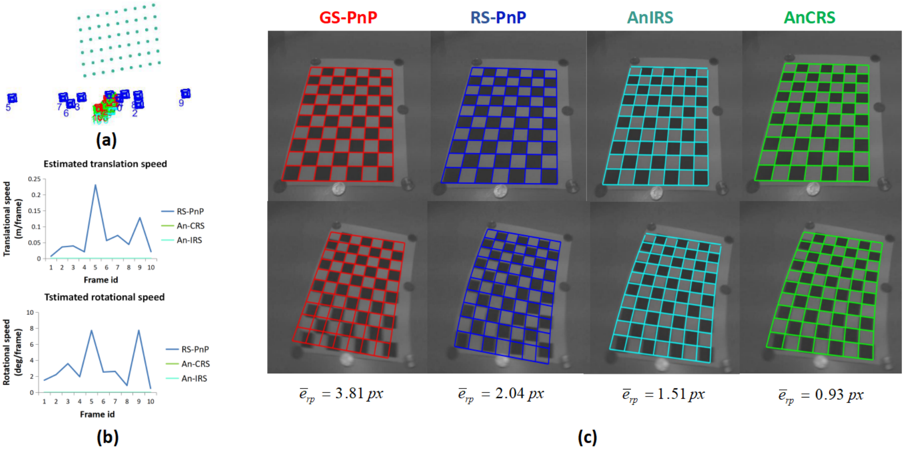

Fig. 8. Visual comparison of reprojected object boundaries by different camera pose and ego-motion estimates. e_{rp} is the reprojection error of the 3D marker points.

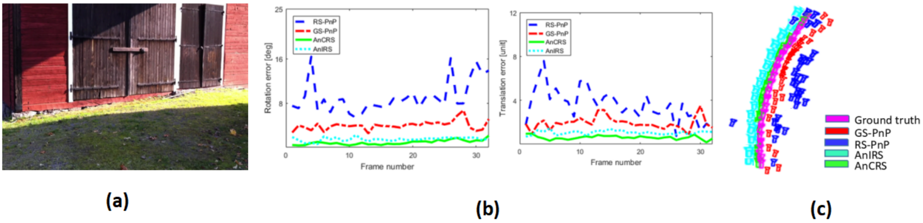

Fig. 9. Results of pose registration with real RS video: **(a)** An example of input RS image. **(b)** Rotation and translation errors of each frame. **(c)** Estimated trajectories by **GS-PnP**, **RS-PnP**, **AnIRS** and **AnCRS** compared to ground truth.

present on the recorded video due to the quick arbitrary camera ego-motion. This scenario can occur in many AR applications. After obtaining the camera pose and ego-motion, the boundaries of the calibration board were reprojected into the RS image. As shown in Fig. 8(c), if the poses and ego-motions are accurately recovered, the reprojected matrix boundaries can perfectly fit the planar marker. In addition to visual checking, the mean value of reprojection errors of 3D marker points of each frame were used as a quantitative measurement.

In the first frame, all four methods obtained acceptable reprojected matrix boundaries due to the small RS effects. However, finding more inliers does not ensure retrieving the true pose and ego-motion, as **RS-PnP** yields 20 geometrically feasible solutions and it is challenging to pick the true one. For example, Fig. 8(a) shows the estimated pose in our AR dataset, where only static camera frames (without ego-motion) were picked. Fig. 8(b) shows that R6P gives distributed locations and huge ego-motion up to 5m/frame, while P3P and our method give similar poses.

In the second frame, with the camera quickly moving, **RS-PnP** and The GS-based method **GS-PnP** provide unstable estimates of camera pose. In contrast, both proposed methods **AnIRS** and **AnCRS** significantly outperform **GS-PnP** and **RS-PnP**. It is noteworthy that **AnCRS** achieves slightly smaller reprojection errors than **AnIRS**. This coincides with the observations made in the synthetic experiments and confirms the theoretical analysis of section 4.2 that the conformal constraint is more suitable to explain the equivalent RS deformations.

Pose registration with real RS video. We tested the four methods for pose registration of an SfM reconstruction. The public dataset [7] was used, which was captured by both RS and GS cameras installed on a rig. The 3D points were obtained by performing SfM with the GS images. 3D-2D correspondences can be obtained by matching RS images to GS images. The results are presented in Fig. 9. The proposed methods **AnIRS** and **AnCRS** give clearly more accurate estimates than **GS-PnP** and **RS-PnP** for most frames.

Running time. The experiments were conducted on an i5 CPU at 2.8GHz with 4G RAM. On average, it took around 2.8s for **AnIRS** (0.1s for isometric reconstruction and 2.7s for 3D-3D registration) and 14.6s for **AnCRS** (10.6s for conformal reconstruction and 4s for 3D-3D registration). Since the proposed method was implemented in MATLAB, an improvement can be expected when using C++ and GPU acceleration, as shown in [27].

7 Conclusion

We have proposed a novel method which addresses the RS-PEnP problem from a new angle: using SfT. By analyzing the link between the SfT and RS-PEnP problems we have shown that RS effects can be explained by the GS projection of a virtually deformed shape. As a result the RS-PEnP problem is transformed into a 3D-3D registration problem. Experimental results have shown that the proposed methods outperform existing RS-PEnP techniques in terms of accuracy and stability. We interpret this improved accuracy as the result of transforming the problem from a 3D-2D registration into a 3D-3D registration problem. This has enabled us to use 3D point-distances instead of the re-projection errors, which carry more physical meaning and make the error terms homogeneous. A possible extension of our work is to derive the exact differential properties of equivalent RS deformation.

Acknowledgement. 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. This research has received funding from the EUs FP7 through the ERC research grant 307483 FLEXABLE.

References

1. Ait-Aider, O., Andreff, N., Lavest, J.M., Martinet, P.: Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In: European Conference on Computer Vision, Springer (2006) 56–68
2. Ait-Aider, O., Bartoli, A., Andreff, N.: Kinematics from lines in a single rolling shutter image. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2007) 1–6
3. Ait-Aider, O., Berry, F.: Structure and kinematics triangulation with a rolling shutter stereo rig. Proceedings of the IEEE International Conference on Computer Vision (Iccv) (2009) 1835–1840
4. Saurer, O., Pollefeys, M., Hee Lee, G.: Sparse to dense 3d reconstruction from rolling shutter images. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2016)
5. Saurer, O., Koser, K., Bouguet, J.Y., Pollefeys, M.: Rolling Shutter Stereo. Computer Vision (ICCV), 2013 IEEE International Conference on (2013)
6. Hedborg, J., Ringaby, E., Forssén, P.E., Felsberg, M.: Structure and motion estimation from rolling shutter video. In: Computer Vision Workshops (ICCV Workshops). (2011)
7. Hedborg, J., Forssen, P.E., Felsberg, M., Ringaby, E.: Rolling shutter bundle adjustment. In: Computer Vision and Pattern Recognition (CVPR). (2012)
8. Dai, Y., Li, H., Kneip, L.: Rolling shutter camera relative pose: generalized epipolar geometry. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4132–4140
9. Kim, J.H., Cadena, C., Reid, I.: Direct semi-dense slam for rolling shutter cameras. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). (May 2016) 1308–1315
10. Albl, C., Sugimoto, A., Pajdla, T.: Degeneracies in rolling shutter sfm. In: European Conference on Computer Vision, Springer (2016) 36–51
11. Ito, E., Okatani, T.: Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on
12. Haralick, R.M., Lee, D., Ottenburg, K., Nolle, M.: Analysis and solutions of the three point perspective pose estimation problem. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE (1991) 592–598
13. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. IEEE transactions on pattern analysis and machine intelligence **25**(8) (2003) 930–943
14. Wu, Y., Hu, Z.: Pnp problem revisited. Journal of Mathematical Imaging and Vision **24**(1) (2006) 131–141
15. Quan, L., Lan, Z.: Linear n-point camera pose determination. IEEE Transactions on pattern analysis and machine intelligence **21**(8) (1999) 774–780
16. Leng, D., Sun, W.: Finding all the solutions of pnp problem. In: Imaging Systems and Techniques, 2009. IST'09. IEEE International Workshop on, IEEE (2009) 348–352
17. Albl, C., Kukelova, Z., Pajdla, T.: R6p-rolling shutter absolute camera pose. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 2292–2300

18. Saurer, O., Pollefeys, M., Lee, G.H.: A minimal solution to the rolling shutter pose estimation problem. In: Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on, IEEE (2015) 1328–1334
19. Albl, C., Kukelova, Z., Pajdla, T.: Rolling shutter absolute pose problem with known vertical direction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 3355–3363
20. Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T., Pizarro, D.: Shape-from-template. IEEE transactions on pattern analysis and machine intelligence **37**(10) (2015) 2099–2118
21. Duchamp, G., Ait-Aider, O., Royer, E., Lavest, J.M.: A rolling shutter compliant method for localisation and reconstruction. In: VISAPP (3). (2015) 277–284
22. Magerand, L., Bartoli, A., Ait-Aider, O., Pizarro, D.: Global optimization of object pose and motion from a single rolling shutter image with automatic 2d-3d matching. In: European Conference on Computer Vision, Springer (2012) 456–469
23. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
24. Magerand, L., Bartoli, A.: A generic rolling shutter camera model and its application to dynamic pose estimation. In: International symposium on 3D data processing, visualization and transmission. (2010)
25. Salzmann, M., Fua, P.: Linear local models for monocular reconstruction of deformable surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(5) (2011) 931–944
26. Brunet, F., Bartoli, A., Hartley, R.I.: Monocular template-based 3d surface reconstruction: Convex inextensible and nonconvex isometric methods. Computer Vision and Image Understanding **125** (2014) 138–154
27. Collins, T., Bartoli, A.: [poster] realtime shape-from-template: System and applications. In: Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on, IEEE (2015) 116–119
28. Chhatkuli, A., Pizarro, D., Bartoli, A., Collins, T.: A stable analytical framework for isometric shape-from-template by surface integration. IEEE transactions on pattern analysis and machine intelligence **39**(5) (2017) 833–850
29. Malti, A., Bartoli, A., Hartley, R.: A linear least-squares solution to elastic shape-from-template. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1629–1637
30. Malti, A., Herzet, C.: Elastic shape-from-template with spatially sparse deforming forces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 3337–3345
31. Haouchine, N., Dequidt, J., Berger, M.O., Cotin, S.: Single view augmentation of 3d elastic objects. In: Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on, IEEE (2014) 229–236
32. Rengarajan, V., Balaji, Y., Rajagopalan, A.: Unrolling the shutter: Cnn to correct motion distortions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 2291–2299
33. Zhuang, B., Cheong, L.F., Lee, G.H.: Rolling-shutter-aware differential sfm and image rectification. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 948–956
34. Haralick, R.M., Lee, D., Ottenburg, K., Nolle, M.: Analysis and solutions of the three point perspective pose estimation problem. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE (1991) 592–598

35. Horn, B.K., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. JOSA A **5**(7) (1988) 1127–1135