Solving Rolling Shutter 3D Vision Problems using Analogies with Non-rigidity

Yizhen Lao, Omar Ait Aider, Adrien Bartoli

Int J Comput Vis manuscript No.
(will be inserted by the editor)

Solving Rolling Shutter 3D Vision Problems using Analogies with Non-rigidity

Yizhen Lao · **Omar Ait-Aider** · **Adrien Bartoli**

Received: date / Accepted: date

Abstract We propose an original approach to absolute pose (AP) and Structure-from-Motion (SfM) which handles Rolling Shutter (RS) effects. Unlike most existing methods which either augment global shutter (GS) projection with velocity parameters or impose continuous time and motion through pose interpolation, we use local differential constraints. These are established by drawing analogies with non-rigid 3D vision techniques, namely Shape-from-Template (SfT) and Non-Rigid SfM (NRSfM).

The proposed idea is to interpret the images of a rigid surface acquired by a moving RS camera as those of a virtually deformed surface taken by a GS camera. These virtually deformed surfaces are first recovered by relaxing the RS constraint using SfT or NRSfM. Then we upgrade the virtually deformed surface to the actual rigid structure and compute the camera pose and ego-motion by reintroducing the RS constraint. This uses a new 3D-3D registration procedure that minimizes a cost function based on the Euclidean 3D point distance. This is more stable and physically meaningful than the reprojection error or the algebraic distance used in previous work. Experimental results obtained with synthetic and real data show that the proposed methods outperform existing ones in terms of accuracy and stability, even in the known critical configurations.

Keywords Rolling Shutter · Absolute Pose · Structure-from-Motion · Non-Rigid · Shape-from-Template

Yizhen Lao
E-mail: lyz91822@gmail.com

Omar Ait-Aider
E-mail: omar.ait-aider@uca.fr

Adrien Bartoli
E-mail: adrien.bartoli@gmail.com

Y. Lao is with College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

O. Ait-Aider and A. Bartoli are with Institut Pascal, Université Clermont Auvergne / CNRS, Clermont-Ferrand, France.

1 Introduction

1.1 Context

Many modern CMOS cameras are equipped with Rolling Shutter (RS) sensors, which are known to be fast, low cost and low power consuming compared to Global Shutter (GS) sensors (El Gamal and Eltoukhy 2005). However, in RS sensors the pixel rows (or columns) are exposed sequentially, e.g. commonly from the top to the bottom of the image. Therefore, the images captured by moving RS cameras are subject to distortions such as wobble and skew, which defeat the classical GS geometric model that is usually assumed in 3D computer vision. In the past decade, many methods have been designed to fit RS camera problems, such as absolute pose (AP) (Ait-Aider et al. 2006, 2007; Ait-Aider and Berry 2009; Saurer et al. 2015), 3D reconstruction from stereo rigs (Ait-Aider and Berry 2009; Saurer et al. 2016, 2013), bundle adjustment for Structure-from-Motion (SfM) (Hedborg et al. 2011, 2012), relative pose estimation (Dai et al. 2016), dense matching (Kim et al. 2016; Saurer et al. 2016) and degeneracy understanding (Albl et al. 2016b; Ito and Okatani 2017). In this paper, we bring a new approach to AP and SfM, two classical and fundamental problems in 3D vision, for the case of RS images. We use RSAP and RSSfM to refer to these problems.

Global shutter absolute pose (GSAP) is the problem of calculating the pose of a calibrated camera with respect to a known 3D model expressed in a world coordinate system.

A special case of GSAP is the so-called PnP problem which consists in computing the pose from n 3D-2D point correspondences. It is important and extensively used in many tasks such as SfM, Simultaneous Localisation And Mapping (SLAM) and Augmented Reality (AR). The general solution for GSAP consists in integrating a minimal problem solver (Haralick et al. 1991; Gao et al. 2003; Wu and Hu 2006; Quan and Lan 1999) in a RANSAC loop (Fischler and Bolles 1981) which both cleans the correspondences and computes the corresponding pose based on a prediction/verification process among putative correspondences. The final step is a non-linear refinement of the pose parameters (Leng and Sun 2009). Obviously, estimating AP in the presence of RS effects with the GS model does not give satisfactory results (Albl et al. 2015, 2019). A few works focus on RSAP (Saurer et al. 2015; Albl et al. 2015, 2016a). They all extend GS-based AP solutions by incorporating the camera motion during image acquisition in the projection model.

SfM aims to recover the 3D scene structure from multiple 2D images with apparent motion. It has been extensively studied for decades. Various applications benefit from it, such as street view mapping and image-based object reconstruction. However, with the intensive use of RS sensors in consumer devices, the RSSfM problem must be considered in real applications, involving, for instance, hand-held cameras, UAV or vehicle embedded cameras.

The RSSfM problem has been studied recently (Hedborg et al. 2012; Albl et al. 2016b; Ito and Okatani 2017; Zhuang et al. 2017; Im et al. 2018). RSSfM takes multiview point correspondences and aims at reconstructing their 3D structure, camera poses and motion. However, all the existing methods impose restrictions on either the movement of the camera (short baseline, smooth motion or pure rotation), the direction of the readout (significant change of the readout direction between views) or the camera model (affine projection). These approaches generally lead to either complex and highly non-linear solutions or use overly restrictive models that limit the application field. Additionally, they are highly sensitive to degenerate configurations which commonly appear in real applications (Ait-Aider and Berry 2009; Albl et al. 2016b; Zhuang et al. 2019).

We present a novel framework to solve RSAP and RSSfM by drawing on recent results obtained in analogous non-rigid reconstruction problems. Specifically, we propose to interpret the RS images of a moving rigid surface as GS images of a virtually deformed surface. By doing so, one can exploit the powerful mathematical formalism and the efficient solutions established in non-rigid vision, namely SfT (for single view deformation estimation of a known template) and NRSfM (for recovering an unknown surface and its deformations from an image set). Having as input a single RS image of a known template or an RS sequence of an unknown surface, the proposed strategy for RSAP and RSSfM has two major steps:

- Step 1: Relaxation. Use either SfT or NRSfM to compute the virtually deformed 3D surface for each image.
- Step 2: Upgrade. Compute the actual pose and (non deformed) structure by reintroducing the RS constraint.

Step 2 treats the pose, structure and kinematics estimation as a purely 3D problem that compares 3D point clouds, the virtually deformed ones and the ones to be recovered.

1.2 Previous Work and Motivation

1.2.1 Previous Work on RSAP

Saurer et al. (2015) propose a minimal solver for RSAP assuming translational motion from 5 3D-2D point correspondences. This solution is limited to specific scenarios, such as a forward moving vehicle. It is not feasible for the majority of applications which depend on a hand-held camera, a drone or a moving robot, where ego-rotation contributes significantly to the RS effect (Hedborg et al. 2012; Duchamp et al. 2015).

Albl et al. (2016a) propose another minimal solver, which also requires 5 3D-2D point correspondences. It is based on a uniform ego-motion model. Nevertheless, it requires the assistance of an inertial measurement unit (IMU), which makes the algorithm dependent on additional sensors. Albl et al. also propose a minimal and non-iterative solution to RSAP called R6P (Albl et al. 2019), which can achieve higher accuracy than the standard P3P (Haralick et al. 1991) by using approximate doubly-linearized (R6P-2lin) or single-linearized (R6P-1lin) models. The approximation used by R6P-2lin requires that the rotation between the 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 estimate from IMU measurements or P3P. This pre-processing step makes R6P-2lin suffer from dependencies on additional sensors or the risk that P3P gives a non satisfactory rough estimate. In contrast, R6P-1lin removes the small rotation assumption and thus is free from the initialization step. Besides, R6P-2lin and R6P-1lin give up to 20 and 64 feasible solutions respectively, which need to be verified, although some of the solutions can be removed by enforcing reasonable values of RS rotational speed. However, they require several hundreds or several thousands of RANSAC iterations, depending on the number of correspondences, to verify all solutions. Experiments showed that both R6P-2lin and R6P-1lin are very sensitive to noise and fail in the case of co-planar points.

Magerand et al. (2012) present a polynomial projection model for RS cameras and propose the constrained global

optimization of its parameters by means of a semidefinite programming problem obtained from the generalized problem of moments method. Contrarily to other methods, this optimization does not require an initialization and can be considered for automatic feature matching in a RANSAC framework. Unfortunately, the method is computationally very expensive.

Oth et al. (2013) propose an RSAP solution for RS calibration which is quite different from the other existing works that augment the GS projection model with the kinematics models. In contrast, they propose to use a high order continuous-time trajectory model combined with the RS model. Thus, both camera pose and shutter time can be recovered by using iterative optimization. However, this solution requires a video sequence as input and uses a frame-by-frame processing which is not able to handle unordered image sets and is also time consuming.

In summary, an efficient and stable solution to RSAP under general motion and without the need for other sensors or restrictive priors is still missing. Such a solution is highly required by many potential applications.

1.2.2 Previous Work on RSSfM

Hedborg et al. (2011); Zhuang et al. (2017); Im et al. (2018) use an RS video sequence to solve RSSfM by assuming smooth camera motion between consecutive frames. The continuous trajectory is estimated by interpolation and specially adapted bundle adjustment. This imposes a high acquisition framerate which results in high computational power requirements. Unordered image sets with large baseline are not handled.

The method in (Ito and Okatani 2017) attempts to solve RSSfM by establishing an equivalence with self-calibrating SfM. The method requires strong priors, namely a pure rotational motion, an affine camera and the availability of one image without RS effects.

Ait-Aider and Berry (2009) first pointed out that pure translation with a velocity vector exactly parallel to the baseline between two camera centres is a case of degeneracy. Lately, Zhuang et al. (2019) further offered a formal proof that RS two-view geometry is degenerate in cases of pure translational camera motion.

A more common degenerate case of RSSfM was pointed out in (Albl et al. 2016b). This work establishes that when the images are taken with similar readout directions, bundle adjustment (BA) with the RS model fails to recover structure and motion. The proposed solution is simply to avoid these degenerate configurations, by taking images with close to perpendicular readout directions. Obviously, this considerably limits the field of use of this method. Note that another approach to avoid the degenerate solution is fusing the information of internal measurement unit (IMU) and video sequence in continuous-time SfM (Lovegrove et al. 2013; Patron-Perez et al. 2015; Ovrén and Forssén 2018, 2019). In the present paper, we focus on RSSfM using exclusively image data and an unordered general set of images with no specific priors.

In summary, a robust and stable solution to solve RSSfM with unordered images and without overly restrictive assumptions on camera motion, readout direction or projection model is still missing. Such a solution would be an important step in the potential widespread deployment of 3D vision with RS imaging systems.

1.3 Contribution and Paper Organization

This paper represents an extension of our previous work (Lao et al. 2018) where we use SfT to solve RSAP. We here extend this principle to RSSfM. Unlike all existing methods which perform 3D-2D registration after augmenting the GS projection model with the velocity parameters, we propose to use local differential constraints. These are established by drawing **analogies** with two non-rigid vision techniques, namely Shape-from-Template (SfT) and non-rigid SfM (NRSfM) (Fig. 1).

Summary of contributions. We have previously shown that the RS effect can be explained by the GS projection of a virtually deformed shape which led to the analogy between the RSAP problem and SfT (Lao et al. 2018). We also proposed a novel RSAP method which first recovers the virtual template deformation using SfT and then computes the pose and ego-motion parameters using a new 3D-3D registration method. In this paper, we not only give a more extensive study of RSAP basing on SfT but also extend the approach to multiple view 3D reconstruction. In summary, the main contributions of this paper are:

- We establish the link between RSSfM and NRSfM by showing that the RS effect in multiple images can be explained as virtual deformations of an unknown surface.
- We propose a novel RSSfM method, illustrated in Fig. 10, which first recovers the virtual deformed structure for each input RS image using NRSfM and then computes the actual rigid structure, camera pose and kinematics using a new 3D-3D registration method.
- Together with our recent conference publication (Lao et al. 2018), we bring a general unified framework to solve RS 3D vision problems which consists in two main steps, namely relaxation and upgrade.

Paper organization. We first introduce the RS projection model and the statement of the RSAP and the RSSfM problems in section 2. We then give a brief introduction to the SfT and NRSfM problems in section 3. The relationships

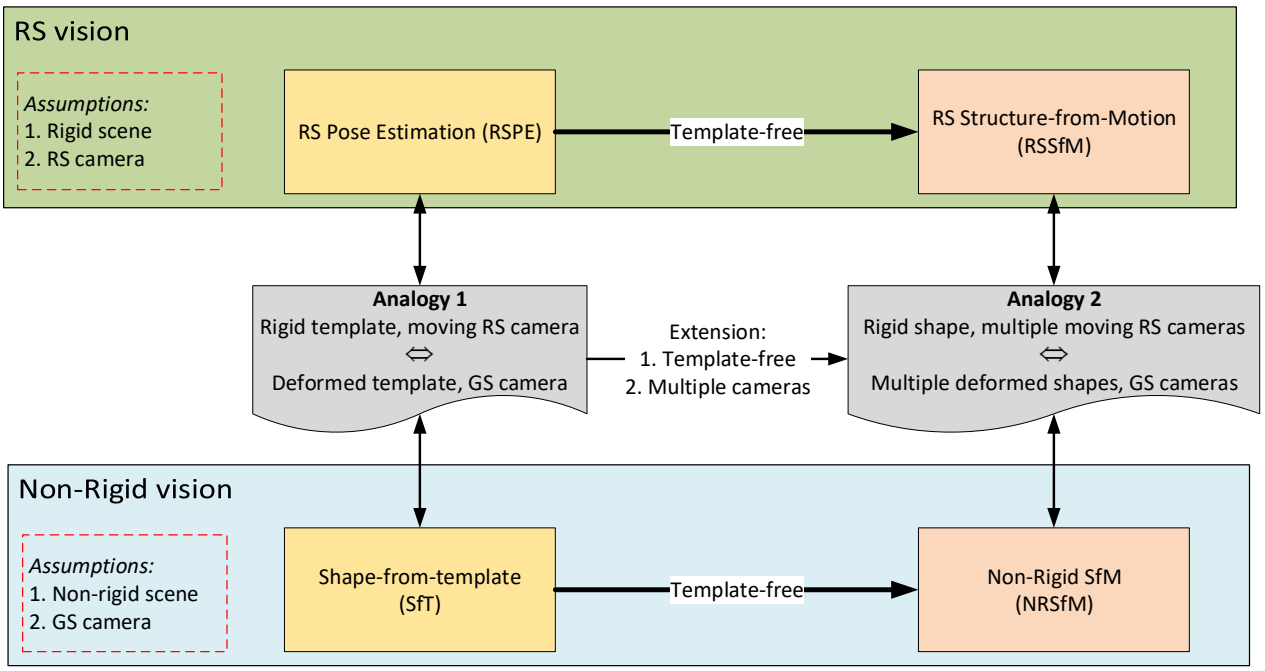

Fig. 1: Overview of the proposed RSAP and RSSfM methods using analogies with non-rigidity.

between RSAP and SfT, and between NRSfM and RSSfM, are analyzed in section 4. We then present a general framework to solve these problems in section 5 followed by two instances applying this principle: in section 6, we show how to solve RSAP by using SfT, while the proposed RSSfM method using NRSfM is presented in section 7. The evaluation of the proposed methods and conclusions are presented in sections 8, 9 and 10.

2 Statement of the Problems

2.1 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 parameter matrix $\mathbf{K}$, rotation $\mathbf{R} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$ between the world and camera coordinate systems (Hartley and Zisserman 2003):

$$\mathbf{q}_i = \Pi^{GS}([\mathbf{R} \quad \mathbf{t}]\,[\mathbf{P}_i^\top \;\; 1]^\top) = \Pi^{GS}(\mathbf{Q}_i) \tag{1}$$

where $\Pi^{GS}([X\ Y\ Z]^\top) = \frac{1}{Z}[X\ Y]^\top$ is the GS projection operator, $\mathbf{P}_i = [X_i\ Y_i\ Z_i]^\top$ is a 3D point in world coordinates, transformed by camera pose to camera coordinates as $\mathbf{Q}_i$. Finally, $\mathbf{q}_i = [u_i\ v_i]^\top$ is its projection in the retina plane, given by normalization from the measured image point $\mathbf{m}_i$ using $\mathbf{K}^{-1}$.

For an RS camera moving during frame exposure, each row is captured in turn and thus with a different pose, yielding a new projection operator Π^{RS}:

$$\begin{aligned}\mathbf{q}_i &= \Pi^{RS}(\mathbf{Q}_i) = \Pi^{GS}(\mathbf{Q}_i^{RS}) \\ &= \Pi^{GS}([\mathbf{R}(v_i) \quad \mathbf{t}(v_i)]\,[\mathbf{P}_i^\top \;\; 1]^\top)\end{aligned} \tag{2}$$

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 The RSAP Problem

With the exception of (Magerand and Bartoli 2010) for RSAP and continuous-time approaches (Lovegrove et al. 2013; Patron-Perez et al. 2015; Ovrén and Forssén 2018, 2019) for RSSfM, most existing methods for RS 3D vision 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 kinematics models are considered in order to express $\mathbf{R}(v_i)$ and $\mathbf{t}(v_i)$. Unfortunately, these additional parameters bring non-linearities in the projection model. A compromise should then be found between the accuracy of the kinematics model and the possibility to

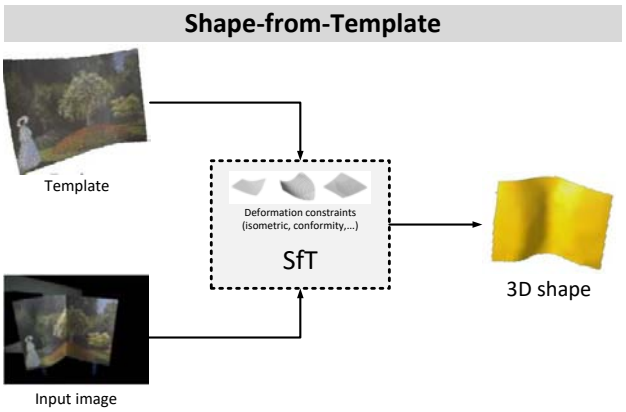

Fig. 2: Illustration of Shape-from-Template (SfT). Example extracted from (Gallardo 2018).

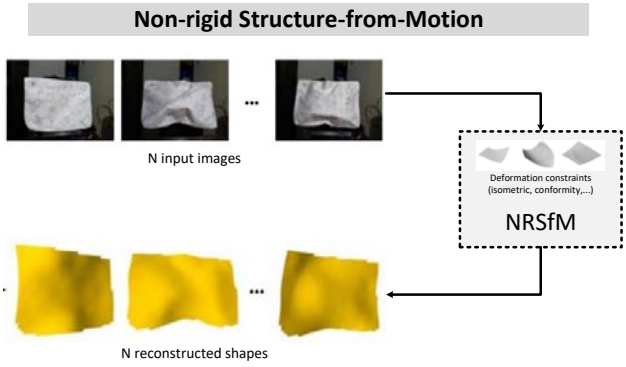

Fig. 3: Illustration of Non-Rigid Structure-from-Motion (NRSfM). Example extracted from (Gallardo 2018).

find an elegant and efficient solution for the RSAP problem. A realistic simplified model is the uniform motion during image acquisition (constant translational and rotational speed). Under this assumption, the RSAP problem consists in computing the camera pose $(\mathbf{R}_0, \mathbf{t}_0)$ corresponding to the first image row and the velocity parameters describing the camera kinematics.

2.3 The RSSfM Problem

The aim of classical rigid SfM is to recover the 3D structure from a set of 2D GS images. Differently, by giving m unordered RS image points $\mathbf{q}_i^j$, the goal of RSSfM is to reconstruct the 3D structure $\mathbf{P}_i$ and to estimate the camera poses $\mathbf{R}_0^j$, $\mathbf{t}_0^j$ as well as the camera kinematics for each of the images $j = 1, \ldots, m$.

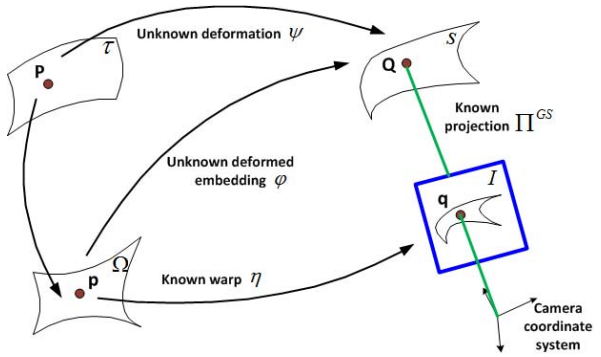

Fig. 4: Geometric model of SfT based on the GS camera.

3 Non-Rigid 3D Vision

We use two techniques designed to address the 3D reconstruction of deformable surfaces: SfT (Fig. 2), which is a template-based approach, and NRSfM (Fig. 3), which estimates the deformations of a surface from a monocular image set.

3.1 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 (Salzmann and Fua 2011; Bartoli et al. 2015). Fig. 4 illustrates a geometric model of SfT. A 3D template $\tau \subset \mathbb{R}^3$ transforms to the deformed shape $S \subset \mathbb{R}^3$ by a 3D deformation $\Psi \in C^1(\tau, \mathbb{R}^3)$. If $\Omega \subset \mathbb{R}^2$ is a 2D space obtained by flattening the 3D template τ, an unknown deformed embedding $\varphi \subset C^1(\Omega, \mathbb{R}^3)$ exists which maps a 2D point $\mathbf{p} \in \Omega$ to $\mathbf{Q} \in S$. Finally, $\mathbf{Q}$ is projected to an image point $\mathbf{q} \in I$ by a known GS projection function Π^{GS}. The known transformation between Ω and I is denoted as η. It is obtained automatically from the 3D-2D point correspondences using Bsplines (Rueckert et al. 1999). 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 in SfT are categorized as follow.

Isometric deformation. The geodesic distances are preserved by the deformation (Bartoli et al. 2015; Salzmann and Fua 2011; Collins and Bartoli 2015; Chhatkuli et al. 2017). 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 may handle isotropic extensible surfaces such as a balloon (Bartoli et al. 2015).

Elastic deformation. Linear (Malti et al. 2015; Malti and Herzet 2017) or non-linear (Haouchine et al. 2014) elastic deformations are used to constrain extensible surfaces. Elastic SfT does not have a local solution, in contrast to isometric and conformal SfT, and requires boundary conditions to be available, such as a set of known 3D surface points.

3.2 Non-Rigid Structure-from-Motion

NRSfM aims to recover the 3D shapes of an object under deformation from a set of 2D GS images. Several NRSfM methods have been presented over the last two decades with various specifications. In particular, (Hu et al. 2013) requires no missing data while (Agudo and Moreno-Noguer 2015; Agudo et al. 2016) require rigid motion at the beginning of the sequence. (Akhter et al. 2009; Gotardo and Martinez 2011) require smooth video sequences. These assumptions do not hold with unordered RS image sets. Besides, some piece-wise methods (Varol et al. 2009; Taylor et al. 2010; Russell et al. 2014) require a segmentation of the image domain into regions, which may be costly with large input datasets, or unavailable. Recently, Kumar et al. (2019) propose a novel NRSfM solution which is able to recovers dense depth without solving for 3D motion parameters. But unfortunately, this approach requires successive frames as input, and does not therefore cope with unordered image set which we focus on in this paper.

4 Analogies Between Rigid RS Projection and Deformable GS Projection

We introduce two analogies between non-rigid vision and RS vision in the single-view and multiple-view cases.

4.1 Template-based, Single-View Case

The main idea is that distortions in RS images caused by camera kinematics 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:

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

In our case the virtual deformation is due to the motion of each surface point during image acquisition. Thus we can denote the deformation as $\Psi^{RS}(\mathbf{P}_i) = \mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i)$. Eq. (3) then becomes similar to Eq. (2):

$$\begin{aligned}\mathbf{q}_i &= \Pi^{GS}(\Psi^{RS}(\mathbf{P}_i)) \\ &= \Pi^{GS}((\mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i))) = \Pi^{RS}(\mathbf{Q}_i)\end{aligned} \tag{4}$$

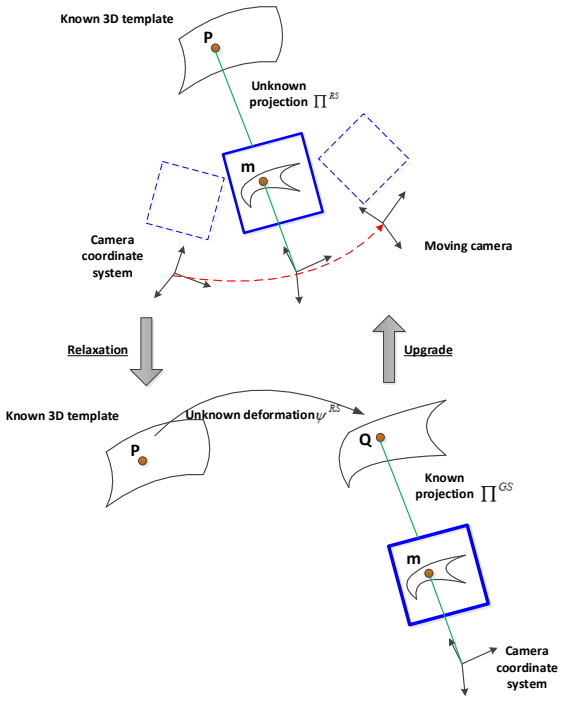

Fig. 5: **Analogy 1**: Equivalence between the RS projection of a rigid object and a GS projection of a virtually deformed object.

Analogy 1: Eq. (4) and Fig. 5 show that the image obtained by a moving RS camera is equivalent to a deformed 3D shape observed 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 Template-free, Multiple-View Case

We now consider an unknown 3D structure observed by a moving RS camera taking multiple images. The analogy described in the previous section can be reused for each image of the sequence.

We define ψ^j as a deformation that maps the original 3D structure $\mathbf{P}_i$ from world coordinates to camera coordinates directly. Then, the RS projection described in Eq. (2) may be rewritten as:

$$\mathbf{q}_i^j = \Pi^{RS}(\mathbf{P}_i^j) = (\Pi^{GS} \circ \psi^j)(\mathbf{P}_i) \tag{5}$$

Analogy 2: Eq. (5) and Fig. 6 show that a set of RS images of a rigid scene may also be interpreted as the same scene under virtual deformations and captured by multiple GS cameras.

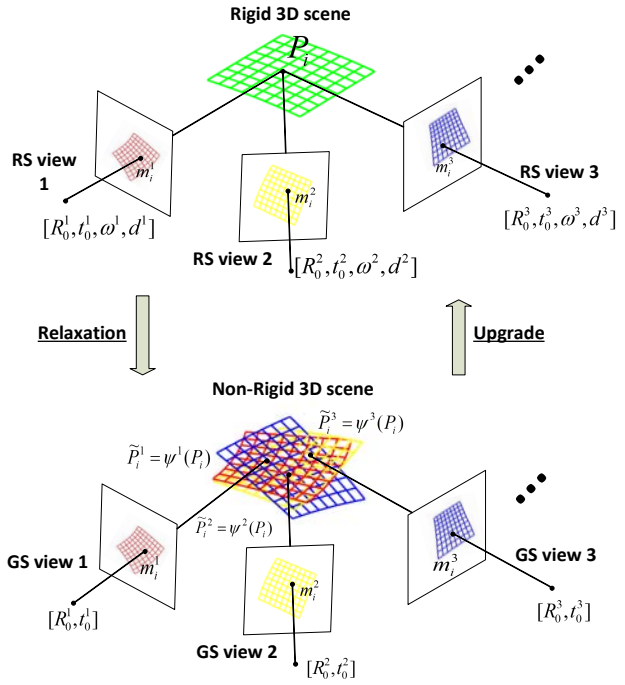

Fig. 6: **Analogy 2**: Equivalence between multiple RS projections of a rigid 3D scene and multiple GS projections of a virtually deformable 3D scene.

Since the deformations are virtual, the 3D scene does not actually deform in the real world. Therefore, we called the original 3D shape $\mathbf{P}_i$ as ***actual structure***, the deformations ψ^j as the ***equivalent RS deformations***, and the virtually deformed shape $\tilde{\mathbf{P}}_i^j = \psi^j(\mathbf{P}_i)$ as the ***equivalent RS deformed shape***.

5 Proposed Solution Framework

The analogies drawn in section 4 allow us to interpret RS images from a new perspective: as GS images of virtual deformations. Thus, in contrast to the existing RS vision solutions which try to constrain the RS projection with various kinematics models, we propose a framework to solve these two problems, which consists in two main steps:

1. **Relaxation.** By interpreting the RS effect as caused by a virtual deformation, we relax the RS constraint of camera kinematics, and transform the problem to NR reconstruction with a GS camera model to recover the equivalent RS deformation.
2. **Upgrade.** We then upgrade the equivalent deformations to the actual rigid structure by reintroducing the RS constraint.

We propose two solutions to RSAP and RSSfM by applying this principle in sections 6 and 7.

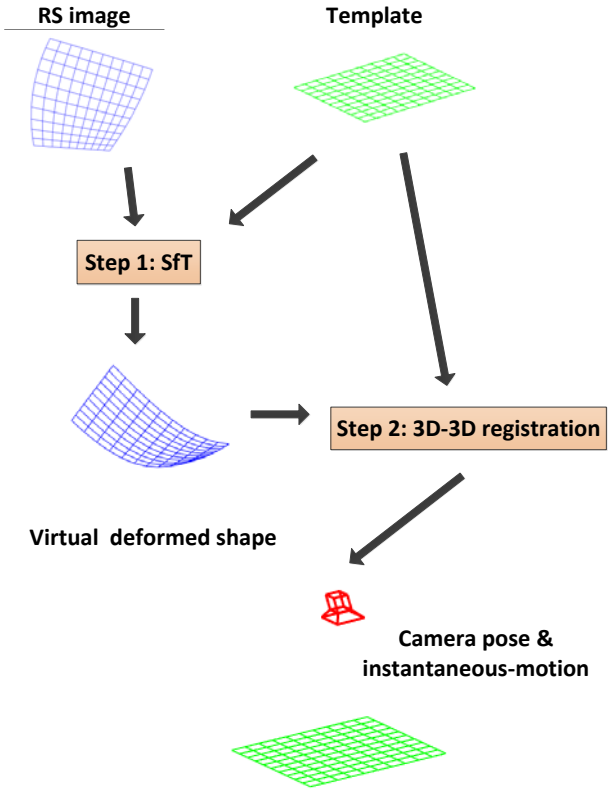

Fig. 7: **An overview of the proposed pose and kinematics estimation method:** *Step 1:* Given an RS image and a known 3D template, we reconstruct the equivalent RS deformed shape using SfT. *Step 2:* By performing 3D-3D registration between the equivalent RS deformed shape and the template, camera pose and kinematics are obtained simultaneously.

6 Solving RSAP using a Virtual Deformation

We introduce the proposed novel RSAP method, illustrated in Fig. 7, which first recovers the virtual template deformation using SfT and then computes the pose and kinematics parameters using 3D-3D registration.

6.1 Step 1: Reconstruction of the Equivalent RS Deformed Shape

After showing the link between the RSAP 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 SfT we should determine which one of the deformation constraints can best describe the equivalent RS deformation.

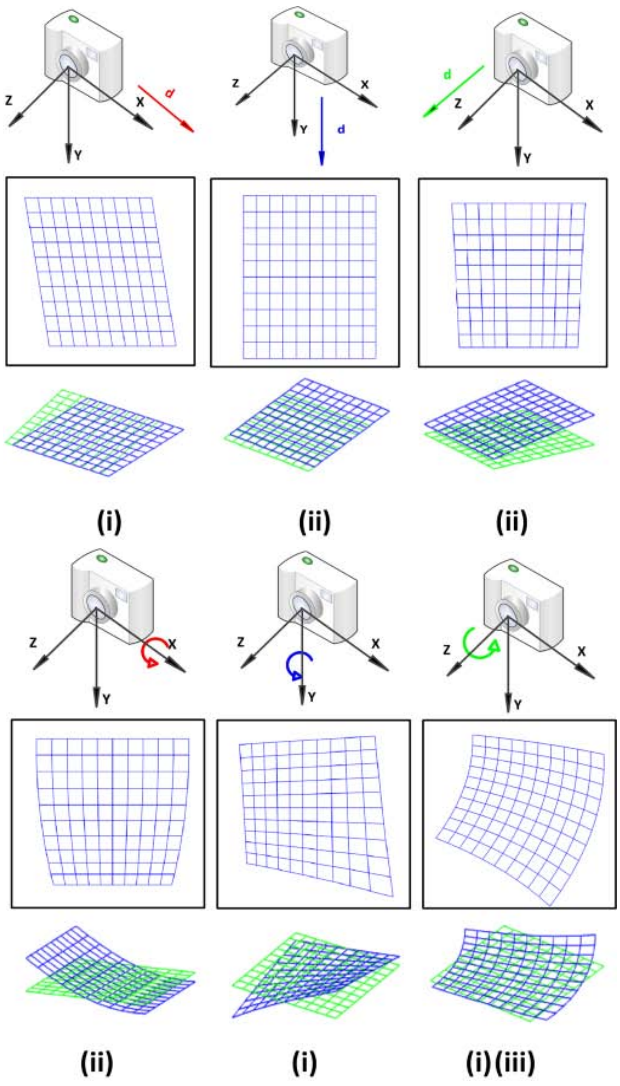

Fig. 8: The 3D template shapes (green shapes in the third and sixth row) captured by an RS camera under different atomic kinematics (first and fourth row) yield distorted RS images (second and fifth row). The exact same images are obtained as the projection of the corresponding virtually deformed 3D shapes (blue shapes in the third and sixth row) into a GS camera. The types of (i), (ii) and (iii) corresponding virtual deformation are given in the main text.

6.1.1 Equivalent RS Deformation under Different Kinematics Models

Any kinematics model can be regarded as a combination of six elementary motions: translation along the X ($\mathbf{d}_x$), Y ($\mathbf{d}_y$), Z ($\mathbf{d}_z$) axes and rotation about the X (ω_x), Y (ω_y), Z (ω_z) axes. Fig. 8 shows RS images and equivalent RS deformed shapes produced by different types of RS kinematics. Albl et al. (2016a) and Rengarajan et al. (2017) illustrated four different types of RS effects in 2D produced by camera motion. Besides, Ovrén et al. (2013) showed the 3D deformations captured by a moving RGB-D camera. In contrast, we base our approach on virtual 3D deformations. Fig. 8 also shows that the corresponding virtual deformations caused by different camera 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.

6.1.2 Choosing the Appropriate Deformation Prior of SfT

It is important to notice that the virtual deformations do not follow any existing 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. The elastic surface may stretch in one direction and generally produces shrinking 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. 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 kinematics, which is the composition of six types of atomic kinematics. Therefore, different surface patches on the shape could be under different 3D deformations. Importantly, the isometric and conformal SfT solutions we used from (Bartoli et al. 2015) exploit **local** differential constraints and recover the local deformation around each point on the shape independently. The assumption we implicitly make is thus not to be taken at the global image level but in the neighbourhood of each point. This turns out to be a very mild and valid assumption in practice.
- The analytical solutions to SfT using the isometric and conformal models reported in (Bartoli et al. 2015), are fast and show the potential to form real-time applications (Collins and Bartoli 2015). In contrast, the existing solutions to the elastic model are slower (Malti et al. 2015; Malti and Herzet 2017) and require boundary conditions unavailable in RSAP .

Isometric deformation. Bartoli et al. (2015) showed that only one solution exists to isometric surface reconstruction from a single view and proposed the first analytical algorithm. A stable solution framework for isometric SfT has been proposed later (Chhatkuli et al. 2017). Thanks to the existing isometric algorithms, we can then stably and efficiently obtain a single reconstruction of the equivalent RS deformed shape $\Psi^{RS}(\mathbf{P}_i)$, $i = 1,\dots,n$.

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 (Bartoli et al. 2015). 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 $\Psi^{RS}_j(\mathbf{P}),\quad j = 1,\dots,M$ are obtained. As shown in Fig. 9 the 2D points located close to each other in the scanning direction in the image are segmented into b groups $\mathbb{G}_k, k = 1,\dots,b$ of N_k points. For this task, we use a region-growing algorithm starting from b seed points. Note that b is a parameter set in advanced according to the size of the image. Differently from the classical region growing approach for image segmentation, which performs growing in the two dimensions based on the similarity between seed and neighbors, our grouping algorithm grows the regions along the vertical axis only. The growing criterion is thus the difference of row index being lower than a threshold d_{max}. The stopping criterion of growing in one direction is that the bound of one region reaches the upper or lower bound of another region. In our experiments, the number of groups b is set as 6 and the threshold d_{max} is experimentally set as a 10% length of the image height.

Then, we calculate a global scale factor s_j of each reconstructed equivalent RS deformed shape to the template by using $s_j = \frac{2}{n(n-1)} \sum_{i,i'=1,i\neq i'}^{n} d_{ii'}/d^j_{ii'}$, where $d_{ii'}$ is the Euclidean distance between 3D points $\mathbf{P}_i$ and $\mathbf{P}_{i'}$ and $d^j_{ii'}$ is the Euclidean distance of the corresponding reconstructed 3D points $\Psi^{RS}_j(\mathbf{P}_i)$ and $\Psi^{RS}_j(\mathbf{P}_{i'})$. We run over $i, i' = 1,\dots,n$ and calculate the average value. Finally, we choose the reconstruction $\Psi^{RS}_j(\mathbf{P})$ with the smallest sum of distance differences along the horizontal direction between each equivalent RS deformed shapes $dx^j_{ii'}$ and known 3D template $dx_{ii'}$ as the best solution:

$$\arg\min_{j\in[1,M]} \sum_{k=1}^{b} \sum_{\substack{i,i'=1,\\ i\neq i'}}^{N_k} |s_j dx^j_{ii'} - dx_{ii'}| \tag{6}$$

6.2 Step 2: Camera Pose and Kinematics Computation

6.2.1 Kinematics Model

Various kinematics models have been used in existing work such as spline interpolation methods (Lovegrove et al. 2013; Patron-Perez et al. 2015; Ovrén and Forssén 2018, 2019), SLERP (Hedborg et al. 2012), Rodrigues formulation (Ait-Aider et al. 2006) for the rotation and the constant speed translation (Zhuang et al. 2017). The proposed 3D-3D RS registration can be easily carried out with any kinematics model. Since the acquisition time of a frame is commonly short, we use a constant velocity model (so-called linearized model) which is a good compromise between accuracy and complexity and is widely used in previous work (Magerand et al. 2012; Dai et al. 2016; Albl et al. 2015, 2016b):

$$\begin{aligned} \mathbf{R}(v_i) &= (\mathbf{I} + [\boldsymbol{\omega}]_\times v_i)\mathbf{R}_0 \\ \mathbf{t}(v_i) &= \mathbf{t}_0 + \mathbf{d}v_i \end{aligned} \tag{7}$$

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. 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 $\boldsymbol{\omega}$.

6.2.2 3D-3D Registration

After obtaining the equivalent RS shape $\Psi^{RS}(\mathbf{P})$, we register the virtually deformed shape to the known 3D template

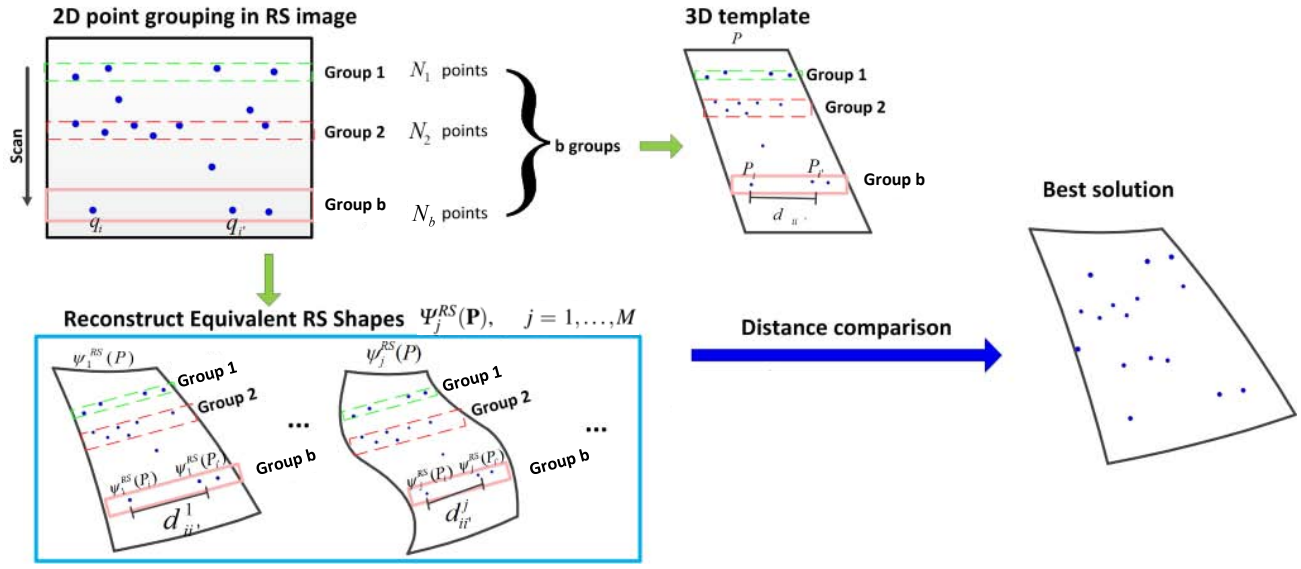

Fig. 9: Choosing the best equivalent RS shape from conformal SfT.

P using the RS kinematics model. By iteratively minimizing the distance errors between the known 3D template and the reconstructed equivalent RS shape using Eq. (7), we can obtain the camera pose and kinematics parameters simultaneously:

$$\underset{\mathbf{R}_0,\mathbf{t}_0,\boldsymbol{\omega},\mathbf{d}}{\arg\min} \sum_{i=1}^{n} \left\| \mathbf{R}(v_i)\mathbf{P}_i + \mathbf{t}(v_i) - \Psi^{RS}(\mathbf{P}_i) \right\| \tag{8}$$

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.

Initialization: we initialize the parameters in Eq. (8) as follows:

- We propose two strategies to initialise $\mathbf{R}_0$ and $\mathbf{t}_0$: i) Computing the absolute orientation between the equivalent RS shape $\Psi^{RS}(\mathbf{P})$ and the known 3D template **P** using (Horn et al. 1988). ii) Performing a classical GSAP method (Haralick et al. 1991) by using the correspondences from the first group (shown in Fig. 9).
- The kinematics parameters ($\boldsymbol{\omega}$,**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 PnP for each set. (2) Initialize **d** and $\boldsymbol{\omega}$ by computing the relative translation and rotation between groups and dividing by the scan time. Alternatively, we can follow a similar procedure by grouping the points of the deformed surface into subsets of close 3D points, which are then registered by computing a rigid body motion (Horn et al. 1988).

However, in many practical situations, it is more convenient and efficient to set the initial values of **d** and $\boldsymbol{\omega}$ to 0, which in our experiments always allowed convergence toward the correct solution.

Refinement: The Levenberg-Marquardt algorithm is used in the non-linear pose and kinematics estimation from Eq. (8).

6.3 Outlier Rejection

Note that outliers in 3D-2D correspondences appear commonly when performing matching in most of real datasets. As reported in (Chhatkuli et al. 2017), SfT will fail if outliers are not removed. Thus, we use the following outlier rejection procedure: 1) The outliers are firstly rejected by using (Pizarro and Bartoli 2012). Therefore, none or very few outliers remain even in challenging datasets. 2) Besides, we further perform a convex L1-minimization (Chhatkuli et al. 2017) in place of the LLS problem in (Dierckx 1993) to reduce the effect of outliers during the reconstructions.

7 Solving RSSfM Using Virtual Deformations

We introduce the proposed RSSfM method, illustrated in Fig. 10, which first recovers the virtual deformed structure for each input RS image using NRSfM and then computes the actual structure, camera pose and kinematics using 3D-3D RS registration.

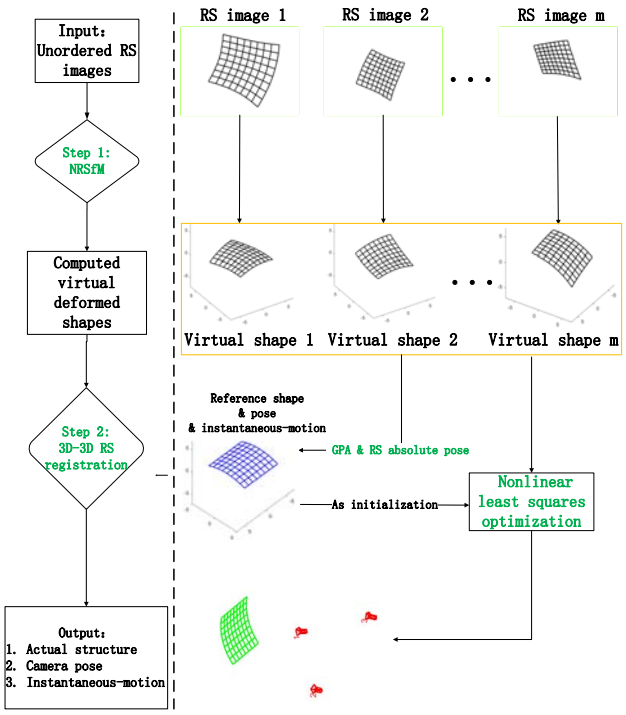

Fig. 10: **Overview of the proposed RSSfM method.** *Step 1:* Given multiple RS images, the equivalent RS deformed shapes are reconstructed using NRSfM. *Step 2:* By performing an iterative 3D-3D RS registration using GPA and RSAP as initialization, the actual structure, camera pose and kinematics are obtained simultaneously.

7.1 Step 1: Reconstruction of the Equivalent RS Deformed Shapes

NRSfM aims to recover the 3D shapes of an object under deformation from a set of 2D GS images. Thus, it allows us to reconstruct the virtual equivalent RS deformed shapes $\tilde{\mathbf{P}}_i^j$ for every RS image.

However, following our discussion in section 3.2, not all NRSfM methods are suitable for RSSfM. We use isometric NRSfM (Iso-NRSfM) (Parashar et al. 2018) for the following reasons:

1. Similarly to SfT and RSAP, isometry is a good approximation to model the equivalent RS deformation.
2. It handles missing data due to occlusions, and unordered input images.
3. It requires $m \geq 3$ views with linear complexity in the number of views and points, and thus combines the use of minimal data with higher efficiency than the other NRSfM methods.

We now briefly describe the two NRSfM methods from (Parashar et al. 2018).

General isometric NRSfM. The Iso-NRSfM method models the object's 3D shape for each image by a Riemannian manifold and deformations as isometric mappings. Each manifold is parameterized by embedding the corresponding retinal plane. This modeling allows one to reason on the metric tensor and Christoffel Symbols (Lee 1997), directly in retinal coordinates, and in relationship to the inter-image warps, which can be computed from point correspondences between images. Based on the theorem that the metric tensor and Christoffel Symbols may be transferred between views using only the warps, a system of two quartics in two variables that involves up to second order derivatives of the warps can be created for an infinitesimally planar surface at each point. An iterative method is then used. The solution of this system are the normals of the surface in all views. The shapes can finally be recovered by integrating the normal field for each view.

Isometric NRSfM with the infinitesimal planarity (IP) assumption. In infinitesimal planarity, one assumes that a surface is at each point locally planar. Thus the surface is globally curved and represented infinitesimally by a set of planes. Since we assume the linearized model for RS kinematics, the virtual equivalent RS deformations are quasi continuous and smooth in the case of wobble, shrinking and extension, which can thus be interpreted by infinitesimal planarity. The general solution for Iso-NRSfM uses the solution with infinitesimal planarity as initialization. However, infinitesimal planarity (InfP-NRSfM) alone gives good results while being even faster than the general algorithm. Therefore, we compare the use of both Iso-NRSfM and InfP-NRSfM to reconstruct the equivalent RS deformed shapes in our experiments.

Discussions of the chosen NRSfM. Note that isometric and conformal SfT are easily and fast solved by existing methods. Their solution is stable and fast to obtain, and they clearly implement two different deformation models, isometry being the strongest one (Bartoli et al. 2015). Differently, NRSfM is a way more difficult problem than SfT, because of the lack of an object model. Two facts are important in our discussion, available from (Parashar et al. 2018): *1)* The first fact is that isometry and conformity actually form the same solution space and methods in NRSfM. In other words, there is no conformal NRSfM. *2)* The second fact is that the equations of isometric NRSfM are tremendously difficult to form and to solve since they depend on the second-order derivatives of the optic flow, in contrast to SfTs, which depend only on the first-order derivatives. Besides, they are nonlinear second-order partial differential equations which currently have no direct solution. The assumption of IP simplifies these equations and allows one to find an initial solution in closed-form by IfRSSfM, which is of key practical

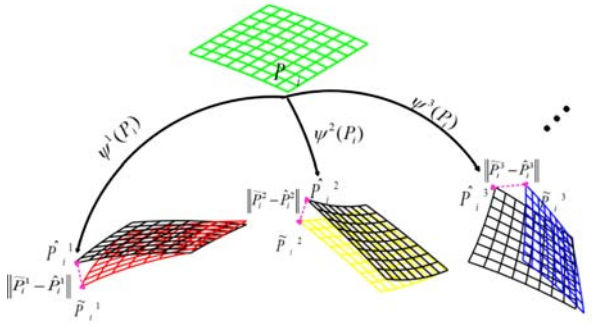

Fig. 11: 3D-3D RS registration recovers the actual shape $\mathbf{P}_i$ (green) by minimizing the sum of squares of the distance between re-deformed shapes (black) $\Psi^j(\mathbf{P}_i)$ and the equivalent RS deformed shapes $\tilde{\mathbf{P}}_i^j$ (red, yellow and blue) recovered by NRSfM.

importance. This solution can then be iteratively refined by exploiting the original equations in IsoRSSfM. It is thus important to understand if the IP solution to NRSfM leads to accurate enough reconstruction estimates in the RS context. In other words, if IfRSSfM can be close to IsoRSSfM in accuracy. Indeed, the IP solution to NRSfM is way faster to compute than the non-IP solution, and could thus be used without the non-IP refinement in a time-critical system, provided that its performance are satisfying.

7.2 Step 2: Recovering the Actual Shape and Cameras

7.2.1 3D-3D RS Registration

After obtaining equivalent RS deformed shapes $\tilde{\mathbf{P}}_i^j$ for each view by NRSfM, we have to estimate the actual shape, camera poses and kinematics. However, the transformations from the actual shape to the equivalent RS deformed shapes are non-rigid. Therefore, as shown in Fig. 11, we design a 3D-3D RS registration by minimizing the sum of squares of the distance difference between the equivalent RS deformed shapes $\tilde{\mathbf{P}}_i^j$ recovered by NRSfM and re-deformed shapes $\Psi^j(\mathbf{P}_i)$ which are obtained from the actual surface under the constraints of the RS kinematics model of each view:

$$\begin{aligned} &\arg\min_{\beta} \sum_{j=1}^{m}\sum_{i=1}^{n} V_i^j \left\| \tilde{\mathbf{P}}_i^j - \Psi^j(\mathbf{P}_i) \right\|^2 \\ &\text{with} \quad \beta = \left\{ \mathbf{P}_i, \mathbf{R}_0^j, \mathbf{t}_0^j, \omega^j, d^j \right\} \\ &\qquad i = 1,\ldots,n, \quad j = 1,\ldots,m \end{aligned} \tag{9}$$

where V_i^j denote the binary variables that equal 1 if a 3D point $\mathbf{P}_i$ is visible in the j^{th} image and 0 otherwise. The deformation function ψ^j is constrained by the RS kinematics model:

$$\psi^j(\mathbf{P}_i) = \mathbf{R}(v_i^j)\mathbf{P}_i + \mathbf{t}(v_i^j) \tag{10}$$

where $\mathbf{R}(v_i^j)$ and $\mathbf{t}(v_i^j)$ are defined by the linearized model described in Eq. (7).

The cost function in Eq. (9) is non-linear least-squares. The availability of a good initial guess for the actual surface points, camera pose and kinematics is thus crucial to ensure convergence toward the solution. This is addressed in the next section.

7.2.2 Initialization

We propose to use GPA and RSAP. GPA solves the problem of registering between multiple observed shape data (Dryden et al. 1998). In this problem, a reference shape which should be as similar as possible to all observed shapes and one global transformation per observed shape are computed. In RSSfM, we assume that the deformations of a given actual point $\mathbf{P}_i$ are random. Thus the actual scene may be chosen as the average shape of all the registered virtual deformed shapes. We can then roughly estimate the actual scene $\mathbf{P}_i$ by performing GPA using the virtual deformed shapes $\tilde{\mathbf{P}}_i^j$ as observed shapes. Then using RSAP from this rough computed actual scene and the RS images, we find the global camera pose $\mathbf{R}_0^j$, $\mathbf{t}_0^j$ and kinematics ω^j, $\mathbf{d}^j$, $j = 1,\ldots,m$ to initialize the optimization in Eq. (9).

7.2.3 Implementation Details

Iso-NRSfM and InfP-NRSfM (Parashar et al. 2018)[1] are both used to reconstruct the equivalent RS deformed shapes. Then we use the stratified GPA method (Bartoli et al. 2013)[2] to initialize the optimization described by Eq. (9), which is eventually conducted using the Levenberg-Marquardt algorithm.

7.2.4 Planar Degeneracy

The combination of NRSfM and the RS constraints makes the proposed two-step method to work well in the common degenerate configurations of RSSfM. An intuitive explanation to this desirable property is as follows. First, NRSfM reconstructs consistent virtually deformed shapes by considering that the viewed surface is locally smooth and differentiable. This is a convenient prior on the scene structure which, though widely applicable, is not used by any other RSSfM method. Once the 3D surfaces are reconstructed for each image, the RS assumption serves to constrain the

[1] http://igt.ip.uca.fr/∼ab/Research/Local-Iso-NRSfM_v1p1.zip
[2] http://igt.ip.uca.fr/∼ab/Research/SGPA_v1p0.tar.gz

pose and kinematics parameters to be compatible with these while the degeneracy was already resolved at the first step.

Specifically, we explain how using the 3D-3D error to recover the scene structure and camera motion instead of the reprojection error allows us to fix the degenerate configuration uncovered in (Albl et al. 2016b). Albl et al. (2016b) stated that any number of RS images with parallel readout directions can be explained by a planar scene undergoing a rotation about the camera x-axis. Bundle adjustment with the linearized RS model always converges toward this trivial solution. However this case is not degenerate for the proposed 3D-3D method. Note that the method of Albl et al. (2016b) focuses on RSSfM with an unordered image set, which is the same as the case we focus on in this paper and different from the BA approach for RS video sequence RSSfM (Hedborg et al. 2012).

We assume without loss of generality that an RS camera has the pose $\mathbf{R}_0 = \mathbf{I}$ and $\mathbf{t}_0 = \begin{bmatrix}0\ 0\ 0\end{bmatrix}^\top$, while the ground-truth of the kinematics is $\boldsymbol{\omega}^{\mathrm{GT}}$ and $\mathbf{d}^{\mathrm{GT}}$. According to Eq. (7) and (10), a 3D point $\mathbf{P}_i^{\mathrm{GT}} = \begin{bmatrix}X\ Y\ Z\end{bmatrix}^\top$ projects as $\mathbf{m}_i = \begin{bmatrix}u_i\ v_i\end{bmatrix}^\top = \Pi^{GS}((\mathbf{I}+[\boldsymbol{\omega}]_\times v_i)\mathbf{P}_i + \mathbf{d}v_i)$. Bundle adjustment minimizes the sum of squares of the reprojection errors (Albl et al. 2016b):

$$\mathbf{e}_i = \mathbf{q}_i - \Pi^{GS}((\mathbf{I}+[\boldsymbol{\omega}]_\times v_i)\mathbf{P}_i + \mathbf{d}v_i) \tag{11}$$

In our method however, the first step using NRSfM does not have degeneracies (Parashar et al. 2018). After obtaining the equivalent deformed shape $\tilde{\mathbf{P}}_i^j = (\mathbf{I}+[\boldsymbol{\omega}^{\mathrm{GT}}]_\times v_i)\mathbf{P}_i^{\mathrm{GT}} + \mathbf{d}^{\mathrm{GT}} v_i$, the second step uses the 3D-3D re-deformation error:

$$\mathbf{e}_i = \tilde{\mathbf{P}}_i - \Psi(\mathbf{P}_i) = \tilde{\mathbf{P}}_i - ((\mathbf{I}+[\boldsymbol{\omega}]_\times v_i)\mathbf{P}_i + \mathbf{d}v_i) \tag{12}$$

Obviously, both Eq. (11) and (12) vanish for the correct configuration $\mathbf{P}_i = \mathbf{P}_i^{\mathrm{GT}}$, $\boldsymbol{\omega} = \boldsymbol{\omega}^{\mathrm{GT}}$, $\mathbf{d} = \mathbf{d}^{\mathrm{GT}}$. However, if we alter the 3D scene and camera to the configuration $\mathbf{P}_i = \begin{bmatrix}X\ 0\ Z\end{bmatrix}^\top$, $\boldsymbol{\omega} = \begin{bmatrix}-1\ 0\ 0\end{bmatrix}^\top$, $\mathbf{d} = \begin{bmatrix}0\ 0\ 0\end{bmatrix}^\top$, Eq. (11) still vanish, while Eq. (12) does not. This means that the RS images could be explained by projecting the 3D scene to the plane $Y = 0$ with the specific kinematics ($\boldsymbol{\omega} = \begin{bmatrix}-1\ 0\ 0\end{bmatrix}^\top$). However, this ambiguity does not occur for the proposed 3D-3D RS registration.

7.3 Outlier Rejection

Similar to the case of RSAP, the proposed RSSfM method which uses NRSfM as the first step also has the risk of failing in the presence of outlier correspondences. Thus, we used a dedicated and efficient outlier rejection strategy (Pizarro and Bartoli 2012) which is based on local surface smoothness and is able to handles large proportions of outliers.

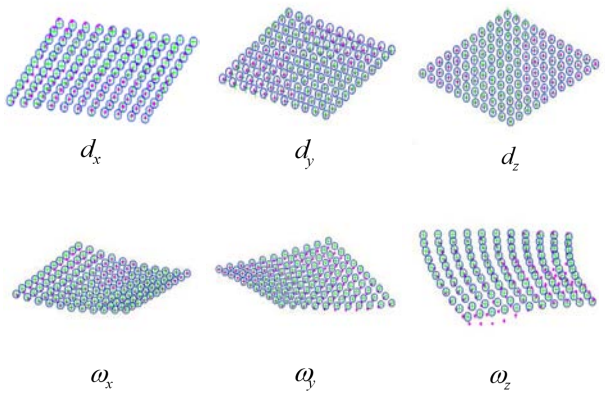

Fig. 12: Reconstructed equivalent RS deformed shapes by **IsoRSAP** (magenta points) and **ConRSAP** (green crosses) compared to ground truth structure (blue circles) under six types of camera kinematics.

8 Experimental Results in RSAP

We compare the proposed methods **IsoRSAP** and **ConRSAP** to two state-of-the-art AP approaches:

- **IsoRSAP:** Our method with the analytical isometric solution to SfT (Chhatkuli et al. 2017)[3].
- **ConRSAP:** Our method with the analytical conformal solution to SfT (Bartoli et al. 2015)[3].
- **PnP**: The GSAP solution (Gao et al. 2003)[4].
- **R6P-2lin**: An RSAP solution (Albl et al. 2015) with double linearized model[5] in RANSAC loop.
- **R6P-1lin**: An RSAP solution (Albl et al. 2019) with single linearized model[6] in RANSAC loop.

8.1 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 cylindrical surface (10 units length and 10 units radius) with an average scanning direction varying from 0 to 90 deg. We 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 $\mathbf{m}$.

[3] http://igt.ip.uca.fr/~ab/Research/SfT_v0p2.zip
[4] estimateWorldCameraPose function in MATLAB
[5] http://cmp.felk.cvut.cz/~alblcene/r6p
[6] https://github.com/CenekAlbl/RnP

8.1.1 Recovering the Equivalent RS Deformed Shape

We first evaluate the ability of **IsoRSAP** and **ConRSAP** to estimate the equivalent RS deformed shapes from RS images. 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 kinematics types (section 6.1.1). For each type, we run 200 trials to obtain statistics. We varied the number of 3D-2D matches from 10 to 121 and used a noise level $\sigma = 1$ px. At each trial, the speed was randomly set with translational speed between 0 and 3 units/frame and rotational speed between 0 and 20 deg/frame.

The results in Fig. 12 show that **ConRSAP** provides stable and high accuracy results for the equivalent RS deformed shape reconstruction while **IsoRSAP** achieves similar performances in the cases of translations and rotation along x-axis. The quantitative evaluation in Table 1 demonstrates that **ConRSAP** generally performs better than **IsoRSAP**. Specifically, it indicates that the advantages of **ConRSAP** are significant in the cases of ego-rotation along the y 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 the other types. Thus, **IsoRSAP** gives slightly better results than **ConRSAP**. All observations confirm our analysis in section 6.1.1 that conformal surfaces can generally better model the equivalent RS deformation.

8.1.2 Pose Estimation

We compared **IsoRSAP** and **ConRSAP** in AP to **PnP**, **R6P-2lin** and **R6P-1lin** with 200 iterations RANSAC. Since the ground truth of camera poses are known, we measured the absolute error of rotation (deg) and translation (units).

Accuracy vs speed. We fixed the number of 3D-2D correspondences to 60 and noise level to $\sigma = 1$ px. We increased the translational speed and rotational speed from 0 to 3 units/frame and 30 deg/frame gradually. At each configuration, we run 100 trials with random velocity directions and measured the average pose errors. The results in Fig. 13 and Fig. 14 show that both **IsoRSAP** and **ConRSAP** provide significantly more accurate estimates of camera orientation and translation with all ego-rotation configurations (ω_x, ω_y and ω_z) compared to **PnP**. **R6P-2lin** achieves better results than **PnP** while the ego-motion speeds are slight. However, with the speed increasing, **R6P-2lin**, which is initialized by **PnP**, losses its accuracy especially when the rotation errors provided by **PnP** is larger than 6 degrees. In contrast, **R6P-1lin**, **IsoRSAP** and **ConRSAP** are not affected and provide stable estimations. Under the six ego-motions, **IsoRSAP** and **ConRSAP** show globally a slight superiority in camera rotation estimation compared to **R6P-1lin**.

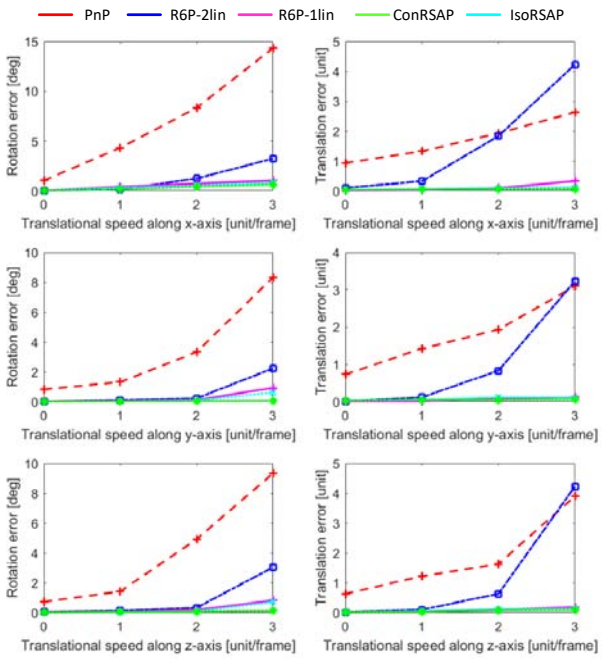

Fig. 13: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different ego-translations.

Accuracy vs image noise. In this experiment, we evaluated the robustness of the five methods against different noise levels. Thus, we fixed the camera translational and rotational speed to 1 unit/frame and 15 deg/frame. Random noise with levels varying from 0 to 2 px was added to the 60 image points. The results in Fig. 15 show that **R6P-1lin**, **IsoRSAP** and **ConRSAP** are robust to the increasing image noise. In contrast, **PnP** is relatively sensitive to image noise. **R6P-2lin** achieves precise estimations with small images noise level (smaller than 2 px). But after **PnP** fails to provide accurate estimation of camera rotation, the accuracy of **R6P-2lin** decreases for both rotation and translation estimation.

Accuracy vs number of correspondences. We evaluated the performance of the proposed methods with different numbers of 3D-2D correspondences. The camera was fixed with translational and rotational speed at 1 unit/frame and 15 deg/frame. The image noise level was set to 1 px. Then we increased the number of correspondences from 10 to 121. The results in Fig. 16 show that the estimation accuracy of all five methods increases with the number of correspondences. **R6P-2lin**, **R6P-1lin**, **IsoRSAP** and **ConRSAP** provide significantly better results in both rotation and translation estimation compared to **PnP**. However, **R6P-2lin** is affected by the inaccurate initializations provided by **PnP** while **R6P-1lin**, **IsoRSAP** and **ConRSAP** are stable.

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 **IsoRSAP** and **ConRSAP** under six types of camera kinematics.

	d_x	d_y	d_z	ω_x	ω_y	ω_z
$\|e_I\|$	0.0130283	0.0113629	**0.0001183**	0.0023273	0.0020031	0.1338190
$\|e_C\|$	**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**

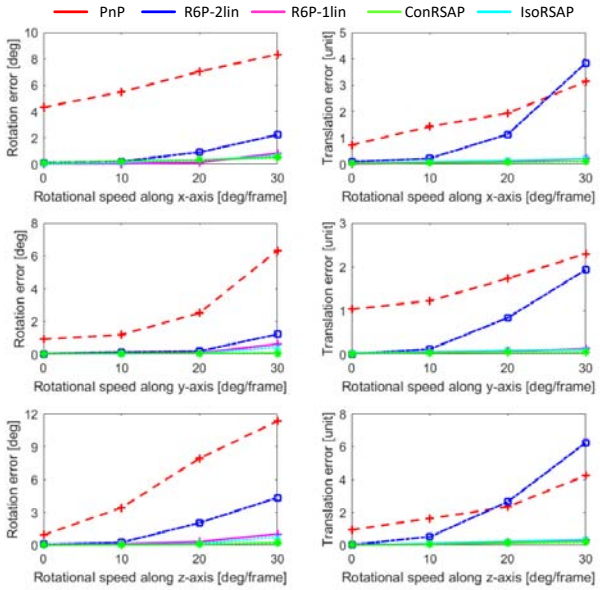

Fig. 14: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different ego-rotations.

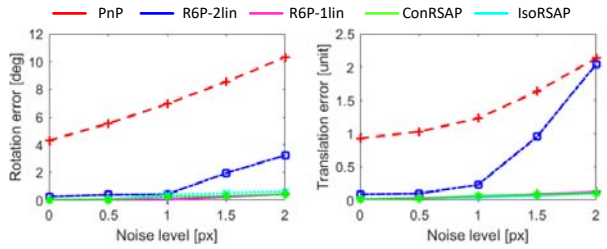

Fig. 15: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different image noise levels.

Accuracy vs outlier rate. In this experiment, we evaluated the performance of the proposed methods against different outlier rates. The number of correspondences is fixed to 60 but with varying outlier rate from 0% to 20%. Following the discussion in section 6.3, **IsoRSAP** and **ConRSAP** perform outliers rejection by using (Pizarro and Bartoli 2012)[7]. The

[7] http://igt.ip.uca.fr/∼ab/Research/FBDSD_v1p0.tar.gz

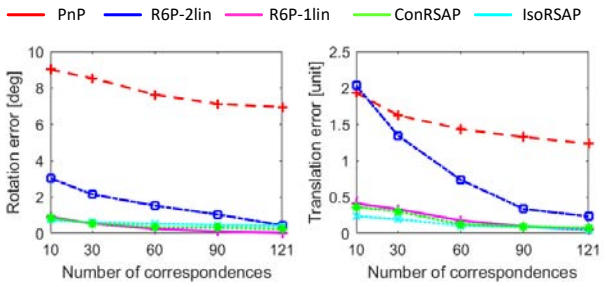

Fig. 16: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different number of correspondences.

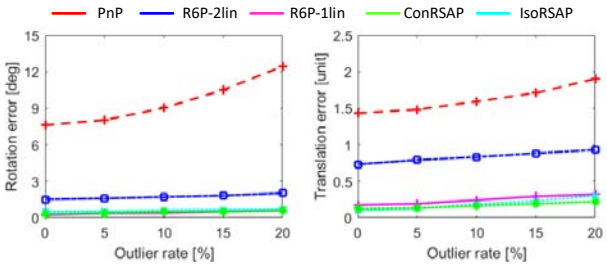

Fig. 17: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different rate of outlier.

results in Fig. 17 show that the estimation error of **PnP** increases significantly with outlier rate. As a result, **R6P-2lin** is increasingly affected by erroneous initializations and also provides slightly increasing estimation errors. In contrast, **R6P-1lin**, **IsoRSAP** and **ConRSAP** show strong robustness against outliers and provide significantly better and stable results.

Accuracy vs curvature. In this experiment, we vary the radius of the surface (inverse of the curvature) from 5 to 30 units. The results in Fig. 18 show that the proposed methods **IsoRSAP** and **ConRSAP** provide stable estimations. The experiment confirms that **R6P-2lin** and **R6P-1lin** do not handle planar or nearly planar scenes with the observation that the estimation errors of **R6P-2lin**, **R6P-1lin** grow rapidly when the inverse curvature is larger than 15 units.

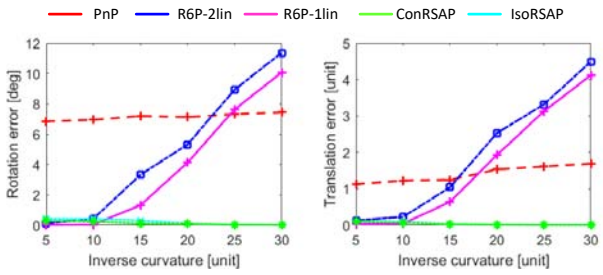

Fig. 18: AP errors for **IsoRSAP**, **ConRSAP**, **PnP**, **R6P-2lin** and **R6P-1lin** under different surface curvatures.

8.2 Real Data

8.2.1 RS Video of a Plane

The five methods have been further evaluated by using real RS images. A chessboard with 64 3D-2D correspondences was captured by a hand-held logitech webcam. Strong RS effects are present on the recorded video due to the quick arbitrary camera motion. After obtaining the camera pose and kinematics, the boundaries of the chessboard were re-projected into the RS image. As shown in Fig. 19, when the poses and velocity are accurately recovered, the reprojected boundaries perfectly fit the chessboard image boundaries. In addition to visual checking, the mean value of reprojection errors of 64 corners of each frame were used as a quantitative measurement.

In the first row of Fig. 19, all methods obtained acceptable reprojected boundaries due to the limited RS effects. However, in the second row, with the camera quickly moving, **R6P-2lin**, **R6P-1lin** and **PnP** provide unstable estimates of camera pose. In contrast, both **IsoRSAP** and **ConRSAP** significantly outperform **PnP** and **R6P**. It is noteworthy that **ConRSAP** achieves slightly smaller reprojection errors than **IsoRSAP**. This coincides with the observations made in the synthetic experiments and confirms the theoretical analysis of section 6.1.1 that the conformal constraint is more suitable to explain the equivalent RS deformations.

8.2.2 RS Video of a Full 3D Scene

We tested the four methods for AP of a 3D scene. The public dataset (Hedborg et al. 2012) 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 are obtained by matching RS images to GS images. Since a large number of correspondences, **R6P-2lin** and **R6P-1lin** run with 1000 iterations of RANSAC. The results are presented in Fig. 20. All the RS methods **R6P-2lin**, **R6P-1lin**, **IsoRSAP** and **ConRSAP** give clearly more accurate estimates than **PnP**. However, we can observe that **R6P-2lin** is affected by **PnP** when the estimates of **PnP** are inaccurate.

8.3 Discussion

From both synthetic and real data experiments, as expected, all RSAP methods **R6P-2lin**, **R6P-1lin**, **IsoRSAP** and **ConRSAP** achieves significantly better estimates compared to the GS based method **PnP**. However, **R6P-2lin** suffers from the large RS effect due to the bad initialisation provided by **PnP**. In contrast, **R6P-2lin**, **IsoRSAP** and **ConRSAP** provide much more stable results.

A key advantage of the proposed methods is that they work for all types of scene geometries, including coplanar points (contrarily to **R6P-1lin** and **R6P-2lin** which are not designed for coplanar points). Planar scenes are common in real applications such as augmented reality and in man-made environments. It is not uncommon, while moving a camera indoor, to end up seeing a wall or a table top, which are considered planar or very nearly planar objects. There is a large body of tracking systems for planar targets. This makes our approach a general solution for real applications.

8.4 Running Time

SfT has already be made very fast in (Collins and Bartoli 2015; Magnenat et al. 2015; Famouri et al. 2018). Since the our previous work (Lao et al. 2018), we have re-implemented our method using a realtime version of SfT. On average, it took around 0.91s for **IsoRSAP** (0.01s for isometric reconstruction and 0.9s for 3D-3D registration) and 11.6s for **ConRSAP** (10.6s for conformal reconstruction and 1.2s for 3D-3D registration). Note that although **ConRSAP** achieves the most accurate estimation. However, it is time-consuming and not suitable for real-time applications. In contrast, **ConRSAP** shows the potential to work in real-time. A possible way is to derive the closed-form solution to RS 3D-3D registration.

9 Experimental Results in RSSfM

In our experiments, the proposed methods were compared to two state-of-the-art techniques:

- **IsoRSSfM:** The proposed method with Iso-NRSfM.
- **IfRSSfM:** The proposed method with InfP-NRSfM.
- **SfM:** An SfM method close to (Wu 2011)[8].

[8] http://mathworks.com/help/vision/examples/structure-from-motion-from-multiple-views.html

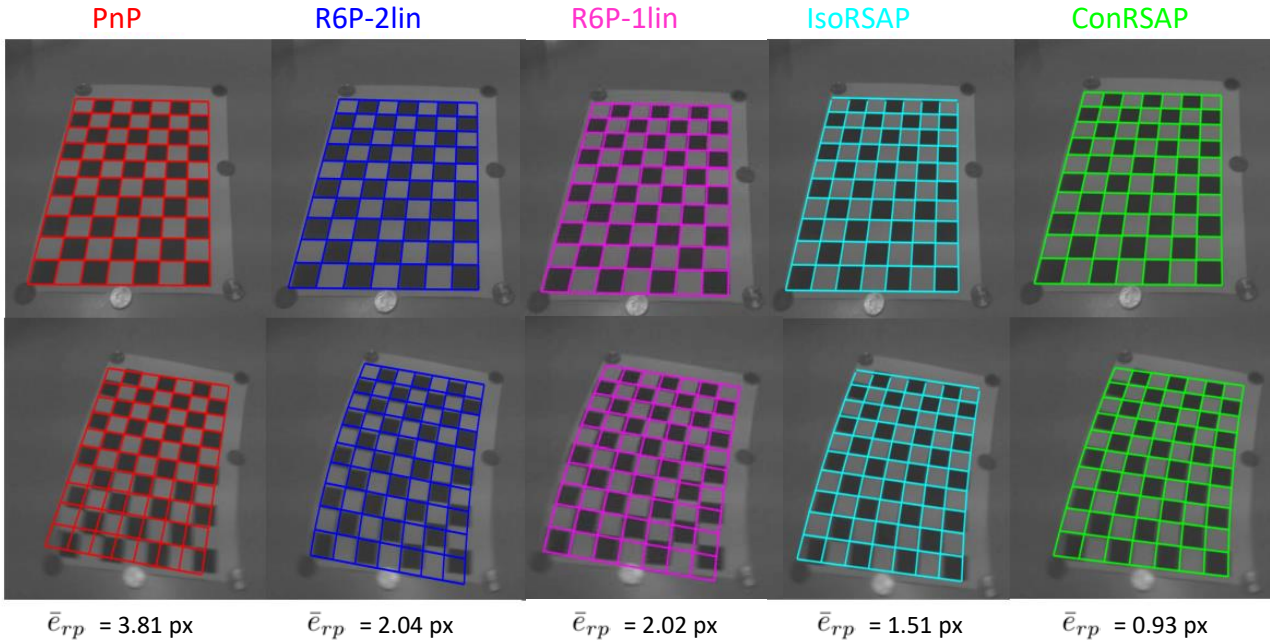

Fig. 19: Visual comparison of reprojected object boundaries by different camera pose and kinematics estimates. $\bar{e}_{rp}$ is the average reprojection error of the 3D marker points.

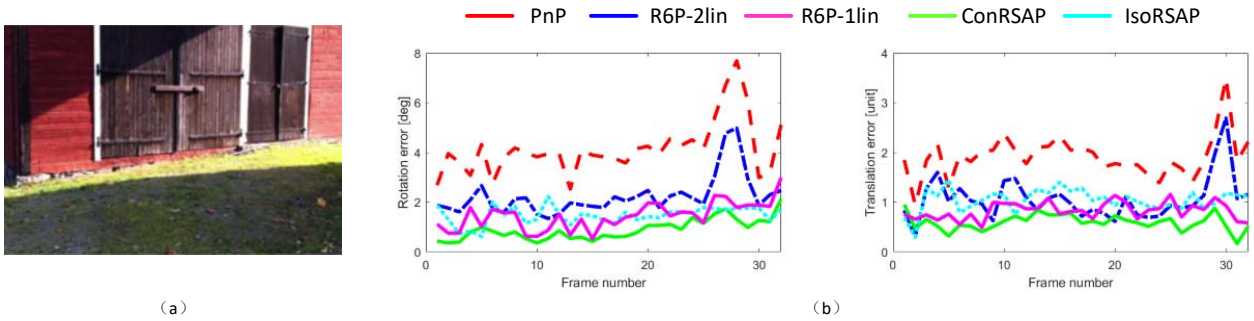

Fig. 20: Results of AP with a real RS video: (a) An example of input RS image. (b) Rotation and translation errors of each frame by **PnP**, **R6P-2lin**, **R6P-1lin**, **IsoRSAP** and **ConRSAP** compared to ground truth (shown in purple).

- **R6PBA:** SfM followed by R6P (Albl et al. 2015) to initialize camera pose and velocity, and refinement by RSBA (Albl et al. 2016b).

9.1 Synthetic Data

We simulated RS cameras located randomly on a sphere with a radius of 20 units and pointing to a cylindrical surface consisting of 81 points. The length of surface is 8 units with a varying radius. The RS image size is 640 px $\times$ 480 px and the focal length 320 px. We compared all methods by varying the speed, the noise on image measurements, the number of views, the surface curvature and the readout direction. The results are obtained after averaging the errors over 50 trials. The default setting is 15 degs/frame and 0.5 units/frame for rotational and translational speed, 1 px noise, 6 views, 15 units radius (inverse curvature).

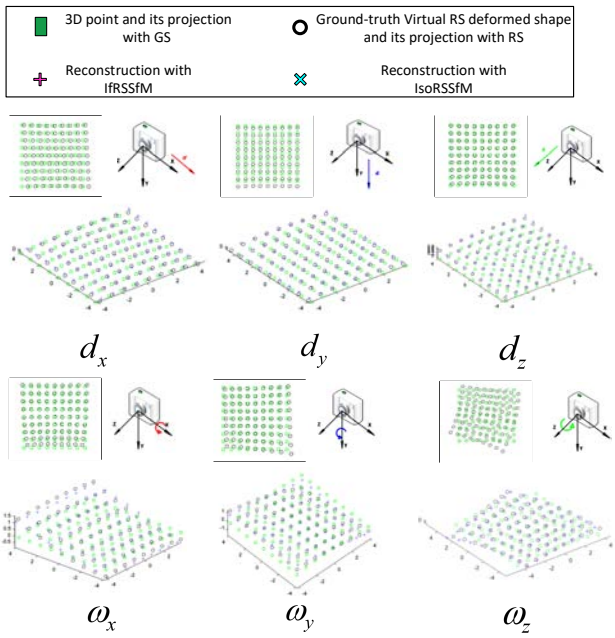

Fig. 21: Deformed shapes reconstructed by **IfRSSfM** and **IsoRSSfM** in comparison to ground truth under six types of camera kinematics.

	d_x	d_y	d_z	ω_x	ω_y	ω_z
$\lvert e_{\text{InfP}}\rvert$	0.067	0.065	0.063	0.115	0.120	0.122
$\lvert e_{\text{Iso}}\rvert$	0.067	0.065	**0.062**	**0.110**	0.120	**0.121**

Table 2: Mean ($\lvert e_{\text{InfP}}\rvert$, $\lvert e_{\text{Iso}}\rvert$) of reconstruction errors (expressed in units) of the equivalent RS deformed shape by **IfRSSfM** and **IsoRSSfM** under six types of camera kinematics.

9.1.1 Reconstructing the Equivalent RS Deformed Shapes

We first evaluate the ability of **IfRSSfM** and **IsoRSSfM** to reconstruct the equivalent RS deformed shapes. We measure the mean distance between the reconstructed 3D points and the corresponding ground truth 3D points computed by Eqs. (5) and (10). The results in Fig. 21 and table 2 show that the two proposed methods accurately reconstruct the deformed shapes under different kinematic types. Although **IsoRSSfM** achieves slightly better reconstruction for $\mathbf{d}_z$, ω_x and ω_z than **IfRSSfM**, no significant visual differences can be observed. This observation verifies the fact that the assumption of infinitesimal planarity is bale to model the globally curved surface with many local infinitesimal planes. Similarly to the discussion of RS deformation in section 6.1.1, isometric surface deformation, which both **IfRSSfM** and **IsoRSSfM** are based on, preserves the distances along all directions, while the equivalent RS distortion only preserves the distances along the horizontal direction. Therefore, we can still observe minor construction errors.

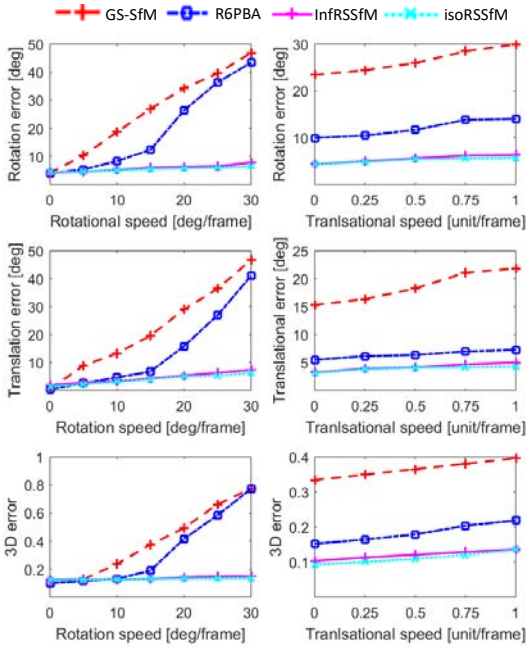

Fig. 22: Camera and shape errors for **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** with increasing rotational and translational speed.

9.1.2 Varying Speed

We evaluated the robustness of the four methods against increasing rotational and translational speed from 0 to 30 degs/frame and 1 units/frame gradually, but with random directions. We measure the reconstruction errors (mean difference between computed and ground truth 3D points in units) and pose errors (mean difference between the computed and ground truth rotation $e_{\text{rot}} = \arccos(\ (\text{tr}(\mathbf{R}\mathbf{R}_{\text{GT}}^{\top}) - 1)/2)$ and translation $e_{\text{trans}} = \arccos(\mathbf{t}^{\top}\mathbf{t}_{\text{GT}}/(\|\mathbf{t}\|\ \|\mathbf{t}_{\text{GT}}\|))$ of each camera in deg). The results in Fig. 22 show that the estimated errors of **SfM** grow with speed. **R6PBA** achieves better results with slow kinematics, while its errors grow dramatically beyond 15 degs/frame. In contrast, both **IfRSSfM** and **IsoRSSfM** provide the best results under all configurations.

9.1.3 Varying Noise Level

In Fig. 23, we observe that the errors for all methods increase linearly when noise varies from 0 to 3 pixels. However, **SfM** shows a better tolerance to noise than **R6PBA**

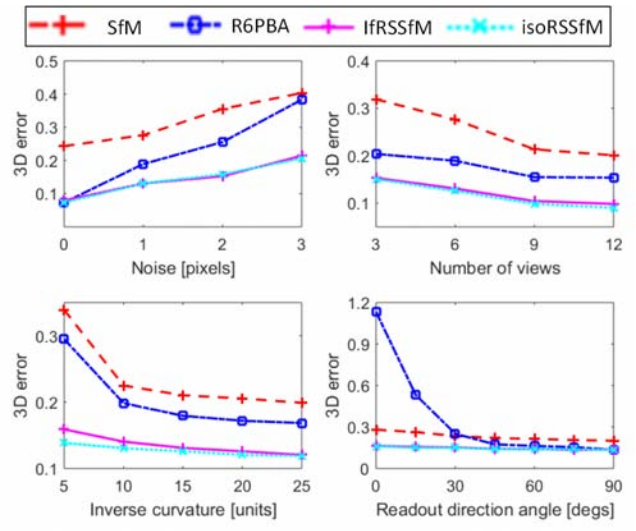

Fig. 23: Reconstruction errors for **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** under different noise levels in image, numbers of views, curvatures and readout directions.

even though its global performance is lower. Both proposed methods achieve the best performance with all noise levels.

9.1.4 Varying Number of Views

Fig. 23 shows that all the four methods give descending errors from 3 to 12 views. **IfRSSfM** and **IsoRSSfM** provide similar results, outperforming **SfM** and **R6PBA**.

9.1.5 Varying Curvature

In this experiment, we vary the radius of the surface (inverse of the curvature) from 5 to 30 units. The results in Fig. 23 show that all the four methods perform better with smaller curvature. The performance of **IfRSSfM** and **IsoRSSfM** are the best among the compared methods. However, as expected **IsoRSSfM** provides slightly better results than **IfRSSfM** when the curvature is large.

9.1.6 Varying Readout Direction

We evaluate the robustness of the four methods with an RS critical motion sequence. We vary the readout directions of the cameras from parallel to perpendicular by increasing the mean angle between them from 0 deg to 90 degs (degenerate to stable). In Fig. 23, we observe that **R6PBA** provides better results than **SfM** with at least 30 deg readout direction. While smaller, the reconstruction error of **R6PBA** grows dramatically, which means that it collapses into the planar degenerate solution. As expected from the analysis in section 7.2.4, **IfRSSfM** and **IsoRSSfM** provide stable results under all settings.

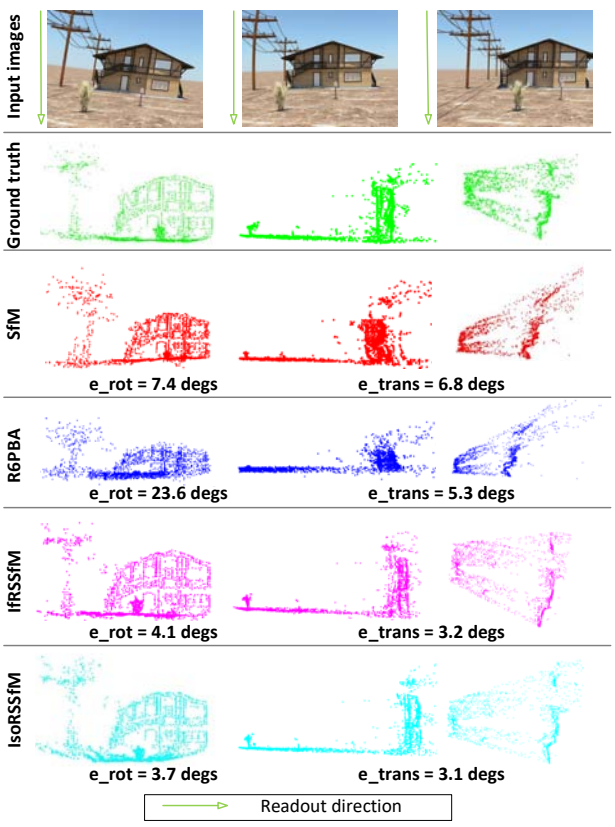

Fig. 24: Reconstruction results and camera errors of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** for synthetic RS images.

9.1.7 Data from Public Benchmark

We tested the four methods on synthetic RS image datasets from (Forssén and Ringaby 2010). We generated unordered image sets by randomly selecting 2 image triplets. In Fig. 24, we observe that quantitatively our methods work better in motion estimation and that qualitatively **SfM** obtains a deformed reconstruction, while **R6PBA** performs worse and provides an extremely deformed reconstruction. In contrast, **IfRSSfM** and **IsoRSSfM** provide reconstructions close to ground truth.

9.2 Real Data

9.2.1 Planar Marker Dataset

We use the RS video previously used in section. 8.2.1 which captures a chessboard with strong RS effects. First, the frames from the video sequence were manually categorized into vertical and horizontal readout direction. Then we designed two kinds of experiments: *1)* We randomly chose 3 images from the vertical group and horizontal group re-

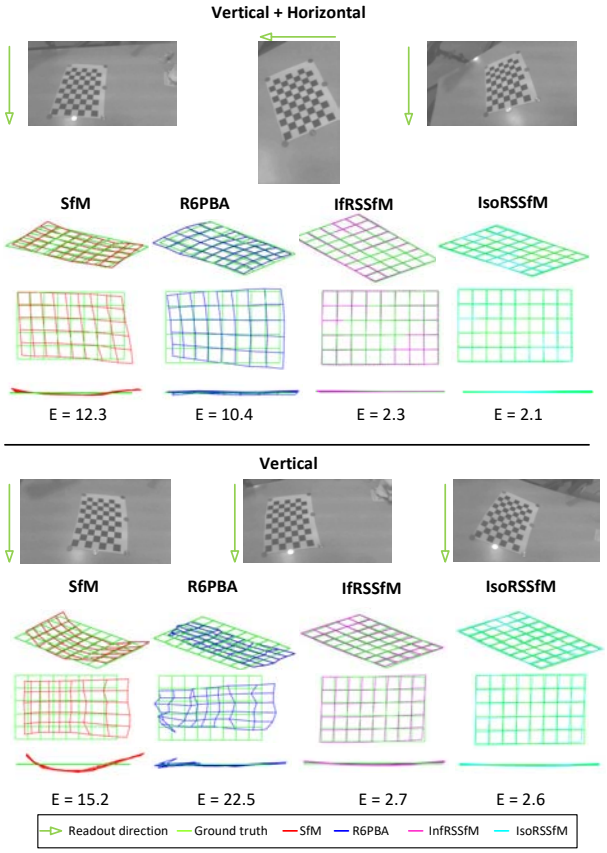

Fig. 25: Reconstructed shapes and mean of reconstruction errors E (in mm) of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** with vertical+horizontal and vertical as inputs respectively for the planar marker dataset.

spectively. *2)* We randomly chose 6 images from the vertical group only. Since the rigid 3D shape is known, we measured the mean distance difference between the computed and ground truth 3D points. The results in Fig. 25 show that **SfM** provides deformed reconstructions in both experiments. **R6PBA** obtains better results than **SfM** in the vertical+horizontal experiment, while it suffers from the planar degeneracy and gives a strongly deformed shape in the vertical-only experiment. In contrast, **IfRSSfM** and **IsoRSSfM** provide a correct reconstruction in both experiments.

9.2.2 Cup and Box Datasets

A cylinder cup and a cubic box were captured by a hand-held Logitech webcam with strong RS effects. The videos were with close readout directions during the acquisition. Again, we randomly chose 6 frames from each video sequence. The ground-truth is now not available. Thus, we use two methods to evaluate the reconstruction results: *1)* Visual checking. *2)* For the cup dataset, we fitted the computed shapes with cylinders by using the *pcfitcylinder* function in MATLAB and measured the fitting errors. For the box dataset, we segmented and fitted the computed scenes with three planes in CloudCompare[9]. Thus, the mean value of fitting errors and between normal vector of the three planes (supposed to be 90 degs) are used as quantitative evaluation criteria. We can observe in Fig. 26 and Fig. 27 that **SfM** fails in handling the RS effects and provides deformed reconstructions for the two datasets. Since the readout directions are close to parallel, **R6PBA** obtains extremely deformed results, close to planar. **IfRSSfM** and **IsoRSSfM** perform best in both the visual checking and quantitative evaluations for both datasets.

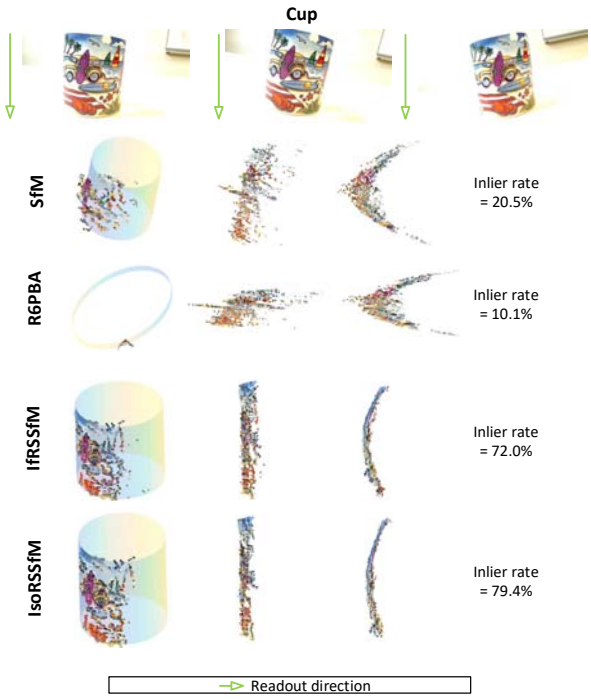

Fig. 26: Visual checking and quantitative evaluations of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** for the cup dataset.

9.2.3 Real RS Sequence

In this experiment, we evaluated the performance of the proposed methods with a challenging real RS sequence (Hedborg et al. 2012) where the camera moves through the scene and the reconstruction grows sequentially. The results in Fig. 28 show that **SfM** is affected by RS effects and provides deformed reconstruction and unsmooth camera trajectory estimation. Again, with parallel readout directions, **R6PBA** provides strongly deformed shape. In contrast, **IfRSSfM**

[9] https://www.danielgm.net/cc/

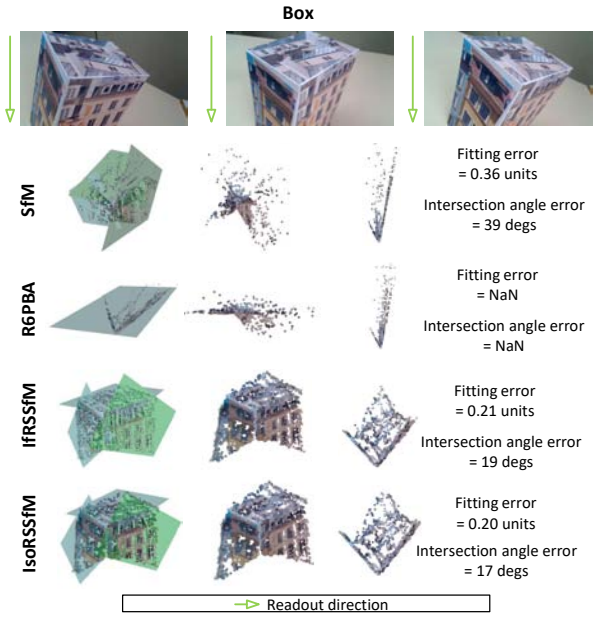

Fig. 27: Visual checking and quantitative evaluations of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** for the box dataset.

Number of points	40	60	80
SfM	4	4	5
R6PBA	12	17	24
IfRSSfM	45	46	49
IsoRSSfM	54	61	67
Number of views	6	9	12
SfM	7	12	20
R6PBA	54	100	153
IfRSSfM	74	93	116
IsoRSSfM	90	109	132

Table 3: Comparison of computation time (in seconds) of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** for 6, 9 and 12 views and 40, 60 and 80 point correspondences with default 3 views and 80 points.

and **IsoRSSfM** provide visually better reconstruction and more reasonable camera pose estimates.

9.3 Running Time

The proposed methods were implemented in MATLAB. The experiments were conducted on an i5 CPU at 2.8GHz with 4G RAM. Table 3 summarises the results and shows that the running time of **SfM**, **IfRSSfM** and **IsoRSSfM** grows slightly with the increasing number of point correspondences and views. In contrast, the computation time of **R6PBA** increases significantly.

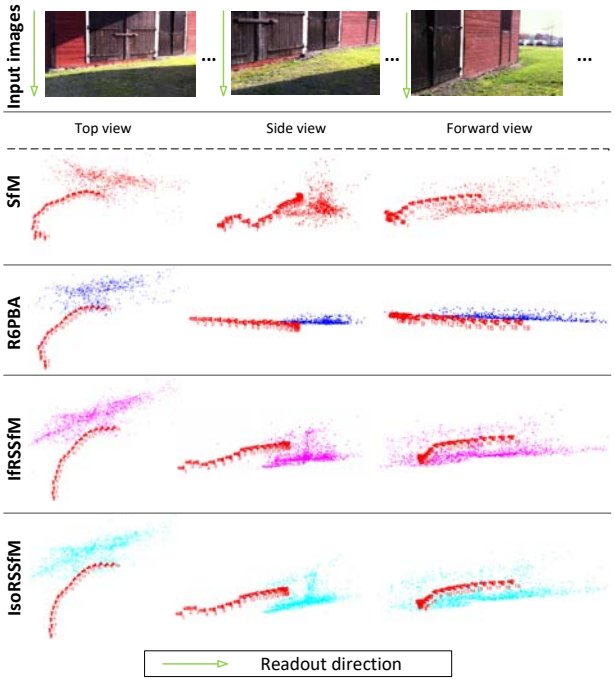

Fig. 28: Visual checking of **SfM**, **R6PBA**, **IfRSSfM** and **IsoRSSfM** for a real RS video.

10 Conclusion

We have presented a novel framework to solve RS vision problems from a new angle. By showing that the RS images of a rigid surface can be interpreted as images of a virtually deformed surface taken by an GS camera, we can first relax the RS constraint and transform the problem to NR reconstruction. Then we upgrade the reconstructed virtual deformations to the actual rigid scene by reintroducing the RS constraints. Based on this framework, we have proposed two novel methods to the RSAP and RSSfM problems respectively.

Firstly, we have proposed novel methods for the RSAP problem using SfT. By analyzing the link between the SfT and RSAP problems we have shown that RS effects can be explained by the GS projection of a virtually deformed shape. As a result the RSAP problem is transformed into a 3D-3D registration problem. Experimental results have shown that the proposed methods outperform existing RSAP techniques in terms of accuracy and stability. We interpret this improved accuracy as the result of two differences compared to existing work: *(i)* By drawing the analogies with non-rigid 3D vision, we solve RSAP locally and analytically. *(ii)* Transforming the problem of 3D-2D registration into 3D-3D registration enables us to use 3D point-distances instead of the re-projection errors, which carry more physical meaning and make the error terms homogeneous. More-

over, the proposed methods work for all types of scene geometries, including coplanar points, contrarily to state of the art methods.

Then we have extented the idea of using non-rigid vision to the RSSfM problem. By showing that the RS effects in multiple images can be explained by multiple virtual deformations of a rigid 3D shape captured by GS cameras, we drew a link between RSSfM and NRSfM. As a result, RSSfM is transformed into a 3D-3D registration problem, which we have shown theoretically and experimentally can successfully avoid the risk of collapsing into a degenerate solution with the usual camera capture manner (parallel readout directions). We have shown that the proposed methods outperform the existing RSSfM methods in accuracy and stability.

Our experiments have also shown that the isometric and conformal deformation models are well suited for the virtual deformations caused by RS effects in most practical applications.

Limitations and perspectives. The observations in our experiments show that the isometric and conformal deformation models well explain the RS equivalent deformation. However, following our discussion in section. 6.1.1, no physics-based constraint in the literature of NR vision can exactly model the RS virtual deformation. This introduces a modelling error for both RSAP and RSSfM. Thus, a possible extension of our work is to derive the exact differential properties of the equivalent RS deformation.

Acknowledgements 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

Agudo A, Moreno-Noguer F (2015) Simultaneous pose and non-rigid shape with particle dynamics. In: CVPR

Agudo A, Moreno-Noguer F, Calvo B, Montiel JMM (2016) Sequential non-rigid structure from motion using physical priors. PAMI

Ait-Aider O, Berry F (2009) Structure and kinematics triangulation with a rolling shutter stereo rig. In: ICCV

Ait-Aider O, Andreff N, Lavest JM, Martinet P (2006) Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In: ECCV

Ait-Aider O, Bartoli A, Andreff N (2007) Kinematics from lines in a single rolling shutter image. In: CVPR

Akhter I, Sheikh Y, Khan S, Kanade T (2009) Nonrigid structure from motion in trajectory space. In: NIPS

Albl C, Kukelova Z, Pajdla T (2015) R6p-rolling shutter absolute camera pose. In: CVPR

Albl C, Kukelova Z, Pajdla T (2016a) Rolling shutter absolute pose problem with known vertical direction. In: CVPR

Albl C, Sugimoto A, Pajdla T (2016b) Degeneracies in rolling shutter sfm. In: ECCV

Albl C, Kukelova Z, Larsson V, Pajdla T (2019) Rolling shutter camera absolute pose. PAMI

Bartoli A, Pizarro D, Loog M (2013) Stratified generalized procrustes analysis. IJCV

Bartoli A, Gérard Y, Chadebecq F, Collins T, Pizarro D (2015) Shape-from-template. PAMI

Chhatkuli A, Pizarro D, Bartoli A, Collins T (2017) A stable analytical framework for isometric shape-from-template by surface integration. PAMI

Collins T, Bartoli A (2015) [poster] realtime shape-from-template: System and applications. In: ISMAR

Dai Y, Li H, Kneip L (2016) Rolling shutter camera relative pose: generalized epipolar geometry. In: CVPR

Dierckx P (1993) Curve and Surface Fitting with Splines. Oxford University Press, Inc.

Dryden IL, Mardia KV, et al. (1998) Statistical shape analysis

Duchamp G, Ait-Aider O, Royer E, Lavest JM (2015) A rolling shutter compliant method for localisation and reconstruction. In: VISAPP

El Gamal A, Eltoukhy H (2005) Cmos image sensors. IEEE Circuits and Devices Magazine

Famouri M, Bartoli A, Azimifar Z (2018) Fast shape-from-template using local features. Mach Vis Appl

Fischler MA, Bolles RC (1981) Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM

Forssén PE, Ringaby E (2010) Rectifying rolling shutter video from hand-held devices. In: CVPR

Gallardo M (2018) Contributions to monocular deformable 3d reconstruction: Curvilinear objects and multiple visual cues. PhD thesis, University Clermont Auvergne

Gao XS, Hou XR, Tang J, Cheng HF (2003) Complete solution classification for the perspective-three-point problem. PAMI

Gotardo PF, Martinez AM (2011) Kernel non-rigid structure from motion. In: ICCV

Haouchine N, Dequidt J, Berger MO, Cotin S (2014) Single view augmentation of 3d elastic objects. In: ISMAR

Haralick RM, Lee D, Ottenburg K, Nolle M (1991) Analysis and solutions of the three point perspective pose estimation problem. In: CVPR

Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge university press

Hedborg J, Ringaby E, Forssén PE, Felsberg M (2011) Structure and motion estimation from rolling shutter video. In: ICCV Workshops

Hedborg J, Forssen PE, Felsberg M, Ringaby E (2012) Rolling shutter bundle adjustment. In: CVPR

Horn BK, Hilden HM, Negahdaripour S (1988) Closed-form solution of absolute orientation using orthonormal matrices. JOSA A

Hu Y, Zhang D, Ye J, Li X, He X (2013) Fast and accurate matrix completion via truncated nuclear norm regularization. PAMI

Im S, Ha H, Choe G, Jeon HG, Joo K, Kweon IS (2018) Accurate 3d reconstruction from small motion clip for rolling shutter cameras. PAMI

Ito E, Okatani T (2017) Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In: CVPR

Kim JH, Cadena C, Reid I (2016) Direct semi-dense slam for rolling shutter cameras. In: ICRA

Kumar S, Ghorakavi RS, Dai Y, Li H (2019) Dense depth estimation of a complex dynamic scene without explicit 3d motion estimation. arXiv preprint arXiv:190203791

Lao Y, Ait-Aider O, Bartoli A (2018) Rolling shutter pose and ego-motion estimation using shape-from-template. In: ECCV

Lee J (1997) Riemannian Manifolds: An Introduction to Curvature. Springer

Leng D, Sun W (2009) Finding all the solutions of pnp problem. In: International Workshop on Imaging Systems and Techniques
Lovegrove S, Patron-Perez A, Sibley G (2013) Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras. In: BMVC
Magerand L, Bartoli A (2010) A generic rolling shutter camera model and its application to dynamic pose estimation. In: International symposium on 3D data processing, visualization and transmission
Magerand L, Bartoli A, Ait-Aider O, Pizarro D (2012) Global optimization of object pose and motion from a single rolling shutter image with automatic 2d-3d matching. In: ECCV
Magnenat S, Ngo DT, Zünd F, Ryffel M, Noris G, Roethlin G, Marra A, Nitti M, Fua P, Gross MH, Sumner RW (2015) Live texturing of augmented reality characters from colored drawings. TVCG
Malti A, Herzet C (2017) Elastic shape-from-template with spatially sparse deforming forces. In: CVPR
Malti A, Bartoli A, Hartley R (2015) A linear least-squares solution to elastic shape-from-template. In: CVPR
Oth L, Furgale P, Kneip L, Siegwart R (2013) Rolling shutter camera calibration. In: CVPR
Ovrén H, Forssén P (2018) Spline error weighting for robust visual-inertial fusion. In: CVPR
Ovrén H, Forssén P (2019) Trajectory representation and landmark projection for continuous-time structure from motion. IJRR
Ovrén H, Forssén P, Törnqvist D (2013) Why would i want a gyroscope on my rgb-d sensor? In: IEEE Workshop on Robot Vision
Parashar S, Pizarro D, Bartoli A (2018) Isometric non-rigid shape-from-motion with riemannian geometry solved in linear time. PAMI
Patron-Perez A, Lovegrove S, Sibley G (2015) A spline-based trajectory representation for sensor fusion and rolling shutter cameras. IJCV
Pizarro D, Bartoli A (2012) Feature-based deformable surface detection with self-occlusion reasoning. IJCV
Quan L, Lan Z (1999) Linear n-point camera pose determination. PAMI
Rengarajan V, Balaji Y, Rajagopalan A (2017) Unrolling the shutter: Cnn to correct motion distortions. In: CVPR
Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast mr images. TMI
Russell C, Yu R, Agapito L (2014) Video pop-up: Monocular 3d reconstruction of dynamic scenes. In: ECCV
Salzmann M, Fua P (2011) Linear local models for monocular reconstruction of deformable surfaces. PAMI
Saurer O, Koser K, Bouguet JY, Pollefeys M (2013) Rolling shutter stereo. In: ICCV
Saurer O, Pollefeys M, Lee GH (2015) A minimal solution to the rolling shutter pose estimation problem. In: IROS
Saurer O, Pollefeys M, Hee Lee G (2016) Sparse to dense 3d reconstruction from rolling shutter images. In: CVPR
Taylor J, Jepson AD, Kutulakos KN (2010) Non-rigid structure from locally-rigid motion. In: CVPR
Varol A, Salzmann M, Tola E, Fua P (2009) Template-free monocular reconstruction of deformable surfaces. In: ICCV
Wu C (2011) Visualsfm: A visual structure from motion system
Wu Y, Hu Z (2006) Pnp problem revisited. Journal of Mathematical Imaging and Vision
Zhuang B, Cheong LF, Lee GH (2017) Rolling-shutter-aware differential sfm and image rectification. In: ICCV
Zhuang B, Tran Q, Ji P, Cheong L, Chandraker M (2019) Learning structure-and-motion-aware rolling shutter correction. In: CVPR