Machine Intelligence Research
www.mi-research.net

20(6), December 2023, 783-798
DOI: 10.1007/s11633-022-1399-z

Rolling Shutter Camera: Modeling, Optimization and Learning

Bin Fan Yuchao Dai Mingyi He

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China

Abstract: Most modern consumer-grade cameras are often equipped with a rolling shutter mechanism, which is becoming increasingly important in computer vision, robotics and autonomous driving applications. However, its temporal-dynamic imaging nature leads to the rolling shutter effect that manifests as geometric distortion. Over the years, researchers have made significant progress in developing tractable rolling shutter models, optimization methods, and learning approaches, aiming to remove geometry distortion and improve visual quality. In this survey, we review the recent advances in rolling shutter cameras from two aspects of motion modeling and deep learning. To the best of our knowledge, this is the first comprehensive survey of rolling shutter cameras. In the part of rolling shutter motion modeling and optimization, the principles of various rolling shutter motion models are elaborated and their typical applications are summarized. Then, the applications of deep learning in rolling shutter based image processing are presented. Finally, we conclude this survey with discussions on future research directions.

Keywords: Rolling shutter, motion modeling, image correction, temporal super-resolution, deep learning.

Citation: B. Fan, Y. Dai, M. He. Rolling shutter camera: Modeling, optimization and learning. *Machine Intelligence Research*, vol.20, no.6, pp.783–798, 2023. http://doi.org/10.1007/s11633-022-1399-z

1 Introduction

A vast majority of prevalent low-budget commercial cameras are built upon complementary metal oxide semiconductor (CMOS) sensors due to their low cost and simplicity in manufacturing, in which a rolling shutter (RS) mechanism is generally employed. Unlike a global shutter (GS) camera capturing all pixels simultaneously using a charge coupled device (CCD) sensor, pixels on the rolling shutter CMOS sensor plane are commonly exposed from top to bottom in a row-by-row fashion with a constant inter-row delay, as illustrated in Fig. 1. Therefore, so-called RS effects (e.g., skew, stretch, and wobble) would occur in the images and videos when relative motion exists between the camera and objects. The RS effect that arises in source media from digital single lens reflex cameras (DSLRs) and other CMOS sensor-based cameras has become a nuisance factor in photography.

Geometric distortions that stem from RS effects violate the conventional perspective camera model[1], thus simply ignoring the RS effect in computer vision applications leads to performance degradation or even failure[2–4]. Considering the rich scene geometry and camera motion information embedded in the row-dependent distortion of RS images[5–8], the high dynamic sampling characteristics of the RS mechanism itself possess underlying spatio-temporal geometric constraints. After nearly 20 years of development, researchers have made full use of the underlying information in the temporal-dynamic imaging mechanism to remove RS distortions and create pleasing visualizations, which has brought great success in downstream applications, such as simultaneous localization and mapping (SLAM)[9, 10] and structure from motion (SfM)[11–14].

Behind these exciting achievements, an in-depth exploration of RS motion models and deep learning techniques is essential. On the one hand, a large number of RS motion models[11, 15] and the resulting non-linear optimization methods[13, 14] have been developed to apply to different types of camera motions, such as discrete, continuous, and special motions. On the other hand, the powerful learning ability of deep learning also empowers various rolling shutter image processing tasks, such as RS image correction[16] and RS temporal super-resolution[5]. Unfortunately, to the best of our knowledge, there is no review to date that systematically investigates the field of RS. Therefore, this paper presents a comprehensive overview of RS cameras for the first time, with a view to promoting the flourishing of this prestigious field.

The remainder of this paper is organized as follows. In Table 1, we show the abbreviations that appear in the text. Fig. 2 presents a general overview of this paper. The

Review
Manuscript received on May 15, 2022; accepted on November 29, 2022
Recommended by Associate Editor Jing-Yi Yu
Colored figures are available in the online version at https://link.springer.com/journal/11633

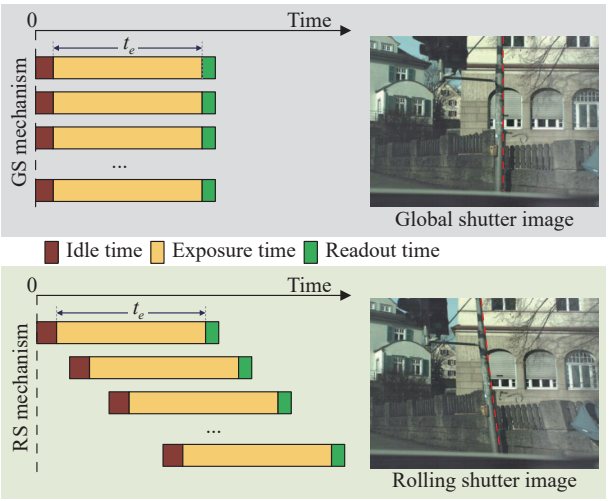


Fig. 1 Illustration of the imaging mechanism of GS cameras (top) and RS cameras (below)

geometric modeling and non-linear optimization methods for RS motion are introduced in Section 2. The typical applications of the RS model are described in Section 3 according to the taxonomy. In Section 4, we review RS image processing methods and datasets, especially in the context of deep learning. Section 5 discusses some future research interests, and Section 6 draws the conclusions.

2 Geometric modeling and non-linear optimization of RS motion

In this section, we summarize the commonly used RS motion models and the corresponding non-linear optimization methods. The discrete model is mainly used for various minimal solver problems (e.g., in relative/absolute pose estimation); the continuous model is more suitable for adjacent frame motion modeling; and the special model can pose better approximations under certain camera motion patterns. Tables 2 and 3 present an overview classification of RS motion models. See Section 3 for their practical applications.

2.1 Discrete motion

Modeling. Due to the temporal-dynamic exposure characteristics of the RS camera, each of its scanlines usually possesses a different projection center, i.e., a series of latent local frames, as shown in Fig. 3. Suppose that the local poses of each scanline of a general RS camera trace out a smooth trajectory in the SE(3) space. In the case of discrete motion, Meingast et al.[15] proposed a seminal geometric motion model in the context of an RS camera, called uniform RS model. With this RS modeling, the smoothly moving camera rotates at a constant angular velocity and translates at a constant linear velocity at the same time. All the projection centers will form a spiral 3D trajectory[2, 17].

As illustrated in Fig. 4(a), we utilize $\boldsymbol{v} \in \mathbf{R}^3$ to express the constant linear velocity and $\boldsymbol{\omega} \in \mathbf{R}^3$ for the constant angular velocity. Note that they describe the inter-scanline translation and rotation displacements, respectively. Note also that here we leverage the axis-angle representation of $\boldsymbol{\omega}$, i.e., $\boldsymbol{\omega} = \omega \boldsymbol{n}$, which can also be represented by spherical linear interpolation (SLERP) in [19, 21, 37] (see Section 2.3 for more instructions). In addition, we assume that the first scanline of the RS image has 6 DoF absolute poses $\boldsymbol{R}_0 \in SO(3)$ and $\boldsymbol{t}_0 \in \mathbf{R}^3$ in the world coordinate system. Formally, the absolute camera poses $\boldsymbol{P}_s = [\boldsymbol{R}_s, \boldsymbol{t}_s]$ of the s-th scanline will satisfy

Table 1 Abbreviation table

Abbreviation	Full name
GS	Global shutter
RS	Rolling shutter
DoF	Degree of freedom
FPS	Frames per second
SfM	Structure from motion
CCD	Charge coupled device
IMU	Inertial measurement unit
SLAM	Simultaneous localization and mapping
RSSR	Rolling shutter temporal super-resolution
CMOS	Complementary metal oxide semiconductor
SLERP	Special linear interpolation
RANSAC	Random sample consensus

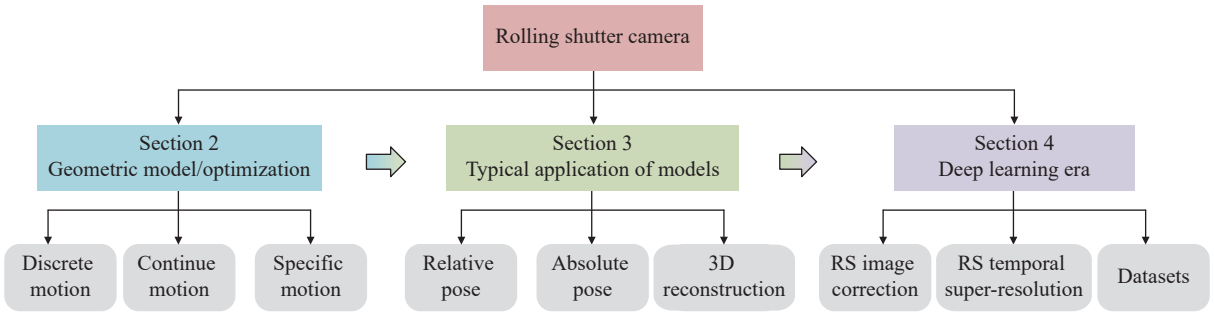


Fig. 2 Overall architecture of this survey. We summarize work in rolling shutter cameras from two aspects of motion modeling and deep learning.

Table 2 Comparison of the uniform RS model as well as its variants[17]. Linear and orbital motions correspond to the pure translation and pure rotation motions in Table 3, respectively.

Motion	Pose $\boldsymbol{P}_s$	Application examples
Linear	$[\boldsymbol{I},\ s\boldsymbol{v}]$	Vehicles traveling in a straight line[17, 18]
Orbital	$[\boldsymbol{I}+s[\boldsymbol{\omega}]_\times,\ \boldsymbol{v}]$	Video clip taken by hand-held devices[19, 20]
Spiral	$[\boldsymbol{I}+s[\boldsymbol{\omega}]_\times,\ s\boldsymbol{v}]$	General RS cameras with smooth motion[3, 4]
Linear	$[\boldsymbol{I}+s[\boldsymbol{\omega}]_\times,\ -s(\boldsymbol{I}+s[\boldsymbol{\omega}]_\times)\boldsymbol{v}]$	3D-2D projection geometry based on continuous video sequences[12, 21]

Table 3 Taxonomy of various well-known RS models and their typical applications

RS model	Motion type	Typical application	Author	Venue	Year	Camera type
Uniform modeling	Full motion	Relative pose	Meingast et al.[15]	Arxiv	2005	Monocular
			Dai et al.[2]	CVPR	2016	Monocular
			Zhuang et al.[6]	CVPR	2019	Monocular
			Lao and Ait-Aider[4]	TPAMI	2020	Monocular
			Wang et al.[22]	ICIP	2020	Stereo
		Absolute pose	Ait-Aider et al.[23]	ECCV	2006	Monocular
			Ait-Aider and Berry[24]	ICCV	2009	Stereo
			Magerand et al.[21]	CVPR	2012	Monocular
			Hedborg et al.[14]	CVPR	2012	Monocular
			Albl et al.[3]	CVPR	2015	Monocular
			Albl et al.[25]	CVPR	2016	Monocular
			Albl et al.[26]	ECCV	2016	Monocular
			Kukelova et al.[27]	ACCV	2018	Monocular
			Albl et al.[28]	TPAMI	2019	Monocular
			Kukelova et al.[29]	ECCV	2020	Monocular
			Albl et al.[30]	CVPR	2020	Stereo
			Wang et al.[31]	RAL	2021	Stereo
	Pure translation	Relative pose	Dai et al.[2]	CVPR	2016	Monocular
			Lao and Ait-Aider[4]	TPAMI	2020	Monocular
		Absolute pose	Saurer et al.[17]	ICCV	2013	Stereo
			Saurer et al.[18]	IROS	2015	Monocular
			Albl et al.[30]	CVPR	2020	Stereo
	Pure rotation	Relative pose	Forssén and Ringaby[19]	CVPR	2010	Monocular
			Ringaby and Forssén[32]	IJCV	2012	Monocular
			Rengarajan et al.[20]	CVPR	2016	Monocular
			Ito and Okatani[33]	CVPR	2017	Monocular
			Lao and Ait-Aider[34]	CVPR	2018	Monocular
			Purkait and Zach[35]	WACV	2018	Monocular
			Lee et al.[36]	Arxiv	2019	Monocular
		Absolute pose	Hedborg et al.[37]	ICCV	2011	Monocular
			Albl et al.[30]	CVPR	2020	Stereo
Continue modeling	Full motion	Relative pose	Zhuang et al.[11]	ICCV	2017	Monocular
			Zhuang and Tran[38]	ECCV	2020	Monocular
		Absolute pose	Im et al.[39]	ICCV	2015	Monocular
			Im et al.[13]	TPAMI	2018	Monocular
			Fan et al.[40]	CVIU	2021	Stereo
			Fan et al.[41]	IVC	2022	Stereo

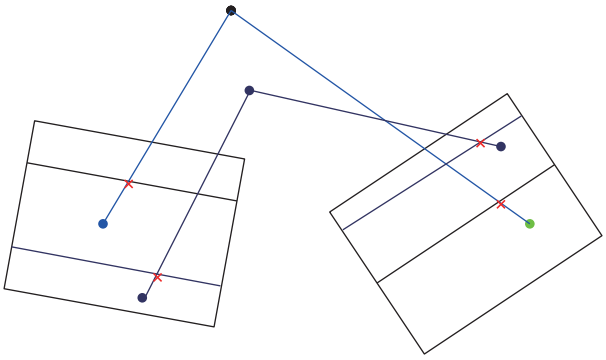


Fig. 3 Illustration of different projection centers of different RS scanlines. For any pair of correspondences (indicated by "×"), the co-planarity constraint still holds [2].

$$\begin{aligned}\boldsymbol{R}_s &= \left(\boldsymbol{I} + \sin(s\omega)[\boldsymbol{n}]_\times + (1-\cos(s\omega))[\boldsymbol{n}]_\times^2\right)\boldsymbol{R}_0\\ \boldsymbol{t}_s &= \boldsymbol{t}_0 + s\boldsymbol{v}\end{aligned} \tag{1}$$

where $[\cdot]_\times$ denotes the skew-symmetric matrix associated with a 3×1 vector. Since the camera typically has a rapid scanning time, it is reasonable to make the assumption that the inter-scanline rotation displacement is sufficiently small. Using the small-rotation approximation, the uniform RS model can be obtained by rewriting (1) as

$$\begin{aligned}\boldsymbol{R}_s &= (\boldsymbol{I} + s[\boldsymbol{\omega}]_\times)\boldsymbol{R}_0\\ \boldsymbol{t}_s &= \boldsymbol{t}_0 + s\boldsymbol{v}.\end{aligned} \tag{2}$$

Optimization. Given N pairs of 3D-2D correspondences, including the 3D point coordinate $\boldsymbol{X}_i \in \mathbf{R}^3$ and the corresponding 2D image coordinate $\boldsymbol{x}_i = (u_i, v_i) \in \mathbf{R}^2$, we can obtain the absolute pose of $\boldsymbol{x}_i$ according to (2) as follows: $\boldsymbol{P}_{u_i} = [\boldsymbol{R}_{u_i}, \boldsymbol{t}_{u_i}]$. Note that $\boldsymbol{X}_i$ is defined in the world coordinate system, which is usually relative to the first scanline of the first frame when processing video images. Consequently, the RS-aware reprojection error can be derived as

$$\boldsymbol{v}^*,\ \boldsymbol{\omega}^* = \arg\min_{\boldsymbol{v},\boldsymbol{\omega}} \sum_{i=1}^{N} \|\boldsymbol{x}_i - \pi(\boldsymbol{X}_i, \boldsymbol{P}_{u_i})\|_2^2 \tag{3}$$

where $\pi(\cdot): \mathbb{P}^3 \to \mathbb{P}^2$ denotes the projection function, defined as

$$\begin{aligned}&\pi(\boldsymbol{X}_i, \boldsymbol{P}_{u_i}) = \langle \boldsymbol{K}(\boldsymbol{R}_{u_i}\boldsymbol{X}_i + \boldsymbol{t}_{u_i})\rangle\\ &\left\langle (x,\ y,\ z)^{\mathrm{T}}\right\rangle = (x/z,\ y/z)^{\mathrm{T}}.\end{aligned} \tag{4}$$

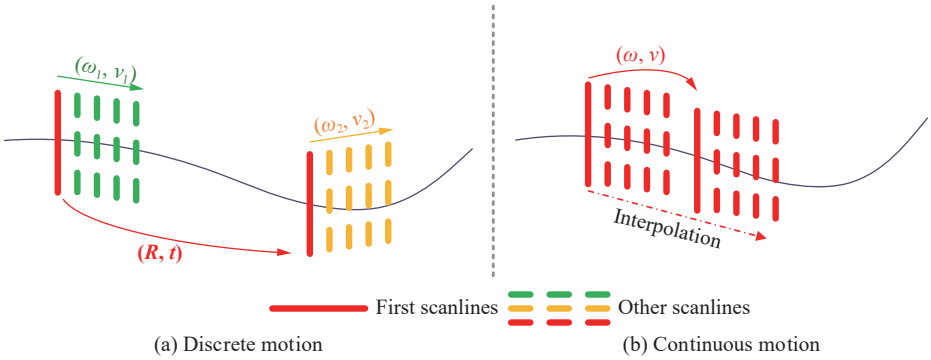
Here, $\boldsymbol{K}$ is the intrinsic matrix whose calibration is easy to implement, e.g., by applying any standard camera calibration procedure to a still scene image captured by a stationary RS camera.

Note that, in (3) we only show the optimization for RS camera motion parameterized by $(\boldsymbol{v}, \boldsymbol{\omega})$. In rolling shutter bundle adjustment (BA)[12–14, 26, 42], it is often necessary to alternatively optimize 3D point coordinates $\boldsymbol{X}_i$ and camera motion $(\boldsymbol{v}, \boldsymbol{\omega})$, similar to the traditional global shutter BA[43]. At this point, (3) will become a standard RS BA problem.

For a moving RS camera, according to (2), each scanline is exposed at a different place in space along the motion trajectory. However, the camera pose is essentially determined by the RS scanline, so the image measurement noise will affect the RS BA due to the coupling between the camera pose and the projected scanline. To solve the non-convex objective (in (3)) more efficiently, most RS BA methods (e.g., [14, 26, 42]) usually ignore the influence of measurement uncertainty on the reprojection function, which may sacrifice the accuracy of RS BA. One way to deal with the measurement noise is error standardisation[44], i.e., by introducing an inverse covariance matrix to provide more confidence for low-variance image measurements. Towards this goal, an acceleration strategy for RS BA is proposed recently in [45], based on the sparsity structure of the Jacobian matrix and Schur complement. Moreover, constructing more accurate correspondences can also alleviate this issue to some extent, for example by filtering outliers through bidirectional consistency check[13, 46].

2.2 Continuous motion

Modeling. Under the continuous camera motion, as depicted in Fig. 4(b), the inter-frame and intra-frame motion smoothness during image acquisition needs to be ex-



(a) Discrete motion (b) Continuous motion

Fig. 4 Illustration of RS motion parameterization, including discrete and continuous camera motions

ploited to account for the scanline-varying camera poses. Specifically, one first assumes that there is a relatively small inter-frame camera velocity $(\boldsymbol{v}, \boldsymbol{\omega})$ between the first two scanlines of two consecutive RS frames (other reference scanlines can be selected without loss of generality). Then, the intra-frame camera motions of all other scanlines can be obtained by interpolation. Formally, the absolute camera position and rotation $(\boldsymbol{p}_1^{s_1}, \boldsymbol{r}_1^{s_1})$ (resp. $(\boldsymbol{p}_2^{s_2}, \boldsymbol{r}_2^{s_2})$) of the s_1-th (respectively, s_2-th) scanline in frame 1 (respectively, frame 2), w.r.t. the first scanline of frame 1 can be expressed as

$$\begin{aligned} \boldsymbol{p}_1^{s_1} &= \lambda_1^{s_1}\boldsymbol{v}, \quad \boldsymbol{r}_1^{s_1} = \lambda_1^{s_1}\boldsymbol{\omega} \\ \boldsymbol{p}_2^{s_2} &= \lambda_2^{s_2}\boldsymbol{v}, \quad \boldsymbol{r}_2^{s_2} = \lambda_2^{s_2}\boldsymbol{\omega} \end{aligned} \tag{5}$$

where $\lambda_1^{s_1}$ and $\lambda_2^{s_2}$ denote the corresponding interpolation factors. Therefore, the relative motion between the s_1-th and s_2-th scanlines obeys the full-motion differential formulation, i.e.,

$$\begin{aligned} \boldsymbol{v}_{s_1 s_2} &= \boldsymbol{p}_2^{s_2} - \boldsymbol{p}_1^{s_1} = (\lambda_2^{s_2} - \lambda_1^{s_1})\,\boldsymbol{v} \\ \boldsymbol{\omega}_{s_1 s_2} &= \boldsymbol{r}_2^{s_2} - \boldsymbol{r}_1^{s_1} = (\lambda_2^{s_2} - \lambda_1^{s_1})\,\boldsymbol{\omega}. \end{aligned} \tag{6}$$

To efficiently model the above interpolation factor, Zhuang et al.[11] proposed a linear interpolation under the assumption of constant velocity motion, i.e.,

$$\begin{aligned} \lambda_1^{s_1} &= \frac{\gamma s_1}{h} \\ \lambda_2^{s_2} &= 1 + \frac{\gamma s_2}{h}. \end{aligned} \tag{7}$$

Here γ is the readout time ratio, which indicates the ratio between the total readout time and the total frame time (including inter-frame idle time) and can be calibrated by [15, 44]. h is the total scanline number in an RS image. In particular, with this RS modeling, the scanline-varying camera poses can be recovered through a simple linear scaling operation. Further and more generally, under the constant acceleration motion assumption, a quadratic interpolation was also proposed by Zhuang et al.[11], i.e.,

$$\begin{aligned} \lambda_1^{s_1} &= \frac{2}{k+2}\left(\frac{\gamma s_1}{h} + \frac{k}{2}\left(\frac{\gamma s_1}{h}\right)^2\right) \\ \lambda_2^{s_2} &= \frac{2}{k+2}\left(1 + \frac{\gamma s_2}{h} + \frac{k}{2}\left(1 + \frac{\gamma s_2}{h}\right)^2\right) \end{aligned} \tag{8}$$

where k denotes the acceleration factor and is in the same direction as the camera velocity, i.e., $k > 0$ for acceleration and $k < 0$ for deceleration. Note that k needs to be estimated additionally when used. Note that it is easy to verify that (8) will reduce to (7) when the acceleration factor vanishes, i.e., $k = 0$.

Optimization. Given N normalized image points $\boldsymbol{x}_i = (x_i, y_i)$, $(i = 1, \cdots, N)$ in scanline s_1^i of frame 1, it corresponds to a forward optical flow of $\boldsymbol{f}^i = (\boldsymbol{f}_u^i, \boldsymbol{f}_v^i)$ and corresponds to a 3D point of depth Z_i. Assume that the scanline where the match point of $\boldsymbol{x}_i$ in frame 2 is s_2^i. Therefore, the vertical optical flow has $\boldsymbol{f}_v^i = s_2^i - s_1^i$. Based on (6), we note that $\beta_i = \lambda_2^{s_2^i} - \lambda_1^{s_1^i}$. Hence, according to the differential formulation[47, 48], the RS-aware differential reprojection error can be developed as

$$\boldsymbol{v}^*,\ \boldsymbol{\omega}^* = \arg\min_{\boldsymbol{v},\boldsymbol{\omega}} \sum_{i=1}^{N} \left\| \boldsymbol{f}_i - \beta_i \left(\frac{\boldsymbol{A}_i \boldsymbol{v}}{Z_i} + \boldsymbol{B}_i \boldsymbol{\omega} \right) \right\|_2^2 \tag{9}$$

where

$$\begin{aligned} \boldsymbol{A}_i &= \begin{bmatrix} -f & 0 & x_i \\ 0 & -f & y_i \end{bmatrix} \\ \boldsymbol{B}_i &= \begin{bmatrix} \frac{x_i y_i}{f} & -\left(f + \frac{x_i^2}{f}\right) & y_i \\ \left(f + \frac{y_i^2}{f}\right) & -\frac{x_i y_i}{f} & -x_i \end{bmatrix} \end{aligned} \tag{10}$$

with f being the camera focal length. More details about RS-aware differential formulation can be found in [11, 40].

2.3 Special motion

In addition to the above-mentioned uniform and differential models for approximating discrete and continuous camera motions, respectively, their simplified versions can often reduce the minimum number of correspondences and the algorithm complexity to compute RS-aware geometry for some specific types of camera motions. For instance, if a car travels along a straight line, the projected center of each scanline will lie on a straight line in 3D space. Therefore, we can naturally ignore the angular velocity $\boldsymbol{\omega}$, i.e., $\boldsymbol{\omega} = \boldsymbol{0}$. At this point, the camera movement becomes a pure translation that is defined only by a constant linear velocity $\boldsymbol{v}$. That is, (2) can be simplified as $\boldsymbol{P}_s = [\boldsymbol{R}_0, \boldsymbol{t}_0 + s\boldsymbol{v}]$, i.e., a linear trajectory is formed. Moreover, except for some wide-angle images, in-plane rotation can be neglected to perform RS correction[49] and RS deblurring[50]. More details on other variants (e.g., orbital motion, linear motion) are provided in Table 2.

On the other hand, other types of motion representations have also received extensive attention. First, without relying on standard kinematic models, the camera trajectory is fitted using a polynomial, e.g., [20, 50, 51]. Afterward, B-splines led to a well-known representation of continuous trajectories in 3D space with better curve-fitting performance. Therefore, many works[9, 44, 52–59] heavily exploit B-spline modeling and parameterize the camera trajectories in terms of 3D knots. As a final note, while the rotation of the camera has predominantly been studied in the case of axis-angles (a.k.a. rotation vectors), more complex representations of rotation are also often explored as follows.

SLERP-based model

It is a linear interpolation on the quaternion sphere, which is used in [14, 19, 37]. It interpolates between two quaternion rotation representations $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, namely,

$$(\boldsymbol{q}_0, \boldsymbol{q}_1, t) = \boldsymbol{q}_0 \frac{\sin(\Omega - t\Omega)}{\sin \Omega} - \boldsymbol{q}_1 \frac{\sin t\Omega}{\sin \Omega}$$

where $\Omega = \arccos\left(\boldsymbol{q}_0^{\mathrm{T}} \boldsymbol{q}_1\right)$. It uses a constant angular velocity in practice and poses a good property (for example, a camera mounted on a rotating platform, which is turning with constant angular velocity). Nevertheless, the presence of sine and cosine yields higher-order variables in the linear polynomial solver[3], resulting in a relatively large computational complexity.

Cayley transform model

For any $\boldsymbol{\eta} = (a, b, c) \in \mathbf{R}^3$, the rotation matrix $\boldsymbol{R}(\boldsymbol{\eta})$ can be represented as

$$\frac{1}{L}\begin{bmatrix} 1+a^2-b^2-c^2 & 2ab-2c & 2b+2ac \\ 2c+2ab & 1-a^2+b^2-c^2 & 2bc-2a \\ 2ac-2b & 2a+2bc & 1-a^2-b^2+c^2 \end{bmatrix}$$

where $L = 1 + a^2 + b^2 + c^2$, which is used to normalize the rotation matrix so that it corresponds to the quaternion of $1 + ai + bj + ck$. It is used for the rotation parameterization of the RS absolute pose problem[28]. Since $\boldsymbol{\eta}$ can also be seen as the axis of rotation scaled by $\tan(\theta/2)$, with θ being the rotation angle, thus the Cayley parameterization is not able to represent the 180-degree camera rotation, resulting in a singular case. More details on interpolation methods for rotations can be found in [60].

3 Geometric problems with RS models

We provide an overview of the applications of the aforementioned geometric RS model from three perspectives: relative pose estimation, absolute pose estimation, and 3D reconstruction. Some representative examples are given in Table 3.

3.1 Relative pose estimation

Relative pose estimation is usually based on a certain number of matched 2D image points and can be used as a subroutine for various computer vision applications, such as visual odometry[61, 62] and SLAM[63, 64].

Dai et al.[2] proposed a generalized RS epipolar constraint for discrete two frames, where 20-pt and 44-pt linear solutions were derived based on the pure translational and uniform models, respectively. Utilizing the consistency of continuous motion between and within frames, Zhuang et al.[11] presented the 8-pt and 9-pt linear solvers based on constant velocity and constant acceleration motion assumptions (see Section 2.2) across two consecutive images, respectively. Fan et al.[40] further extended them to standard RS stereo cameras and proposed an RS-stereo-aware differential epipolar constraint. Wang et al.[22] proposed a linear solution for the standard RS stereo camera using the RS uniform model, requiring only nine correspondences for each of the left and right images. More recently, Lao and Ait-Aider[4] derived an RS-homography matrix to transform two discrete RS images, while Zhuang and Tran[38] proposed a differential RS-homography matrix to compensate for the camera ego-motion between two consecutive frames. Lee et al.[36] proposed a minimal 5-pt solution between two discrete frames with the help of the angular velocity obtained from the gyroscope. Furthermore, simplified motion assumptions detailed in Section 2.3 were considered in [19, 20, 32, 34].

3.2 Absolute pose estimation

In addition to some methods based on known GS templates[8, 54, 65], there is a body of work dedicated specially to the design of linear solvers (e.g., minimal solvers). This is called the RS perspective-n-point (RnP) problem, i.e., estimating the position and orientation of the RS camera based on a certain number of correspondences between 3D scene points and 2D image points.

Ait-Aider et al.[23] presented a closed-form solution based on 8.5-pt by assuming a planar scene. An extension to line correspondences was given in [66]. Magerand et al.[21] developed a globally optimal solution by constructing and minimizing polynomial equations with the Gloptipoly solver, while better results were obtained by at least 7 correspondences compared with the 8.5-pt method. For a forward-moving car, Saurer et al.[18] proposed a minimal 5-pt method using the pure translation model, but this simplified model inevitably limits its application in most real-world applications[4]. Albl et al.[3] initialized the camera pose using the standard P3P method[67], and then utilized the double linearized RS model and the Gröbner basis solver to derive a minimal 6-pt solution (R6P), which requires only six 3D-2D correspondences. Afterward, Albl et al.[28] further solved this double linearized model using hidden variable techniques, obtaining a faster and more accurate solution. Meanwhile, the Cayley parameterization for rotations was studied, removing the reliance on P3P initialization. Later, Albl et al.[25] fused Inertial measurement unit (IMU) data and proposed R5P using known gravity directions. To improve the efficiency of R6P, Kukelova et al.[27] devised a novel alternating iteration strategy, yielding a simpler and faster linear solver. When the camera focal length and radial distortion are unknown, Kukelova et al.[29] further combined this iterative scheme with fast generalized eigenvalue and Gröbner basis techniques to propose a minimal 7-pt solution.

3.3 3D reconstruction

Image-based 3D reconstruction is of great importance in the fields of robot navigation[68], visual perception[69], and 3D modeling[70]. In the following, we summarize the classic 3D reconstruction methods based on RS video sequences or RS stereo images. Note that the RS structure from motion (RS-SfM) is mainly investigated, which inevitably involves absolute pose estimation (from the perspective of non-linear optimization).

Using image pairs taken by two stationary RS cameras, Ait-Aider and Berry[24] used non-linear least-squares to estimate the structure and velocity of rigid moving objects. Saurer et al.[17] proposed a plane sweeping method for RS stereo images, which can solve for the exposure time and the scene depth simultaneously. Subsequently, Saurer et al.[12] developed sparse-to-dense 3D reconstruction for wide-baseline RS images. Hedborg et al.[14] pioneered an RS bundle adjustment method for RS video images by introducing the SLERP model for rotation interpolation. Based on this framework, Im et al.[13] further linearly interpolated the rotation and developed an RS plane sweeping technology for accurate 3D reconstruction. Note that a deep learning version of RS plane sweeping was proposed by Fan et al.[71] Assuming a pure rotation motion, the critical motion sequences of RS-SfM were recast as self-calibration of the imaginary camera in [33]. Additionally, the degeneracies of RS-SfM in multi-view reconstruction were pointed out in [26]. Recently, Albl et al.[30] explored an RS stereo rig, e.g., in smartphones, with a negligibly small baseline and opposite scanning direction, to recover the camera pose and scene geometry. Wang et al.[31] proposed an RS stereo depth estimation method that uses a coarse-to-fine scheme to alternately update the depth map and refine the camera motion. Fan et al.[41] proposed an RS-stereo-aware differential SfM method, compensating for 3D degradation effectively by a simple linear scaling operation.

4 RS image processing

RS image processing has been a hot topic for a long time[5, 11, 16, 38, 72]. Not only RS modeling has contributed to its prosperity, but with the advent of the deep learning era, convolutional neural networks have gradually made a big splash in the field of RS image processing, such as RS image correction, and RS temporal super-resolution. In this section, we will provide an overview of these new paradigms based on deep learning, along with a summary of relevant public datasets.

4.1 RS image correction

RS image correction aims to remove RS artifacts to recover the underlying GS image. Fig. 5 summarizes its development according to a timeline. In the following, for compactness, we first give a brief description of the traditional RS image correction methods. Then, the focus is on the deep learning-based RS image correction approaches.

Traditional methods. Early on, pure translation model[80, 81] and pure rotation model[19, 32, 82] were usually utilized. A representative work comes from Grundmann et al.[82], who proposed a homography mixture to achieve joint RS removal and video stabilization. Geometric information in the scene, e.g., straightness of lines[20] and orthogonal vanishing directions[35] in a Manhattan world has been employed. Lao and Ait-Aider[34] improved over [20, 35] by fitting at least four image curves in a random sample consensus (RANSAC) scheme. The pure rotation model was also used in [20, 34]. Furthermore, some works are dedicated to simultaneously removing RS distortion and handling other image processing/computer vision tasks, such as super-resolution[81, 83], motion deblurring[50, 84], differential SfM[11], image stitching[38], and video stabilization[74]. Very recently, Vasu et al.[73] proposed a multi-layer 3D scene model for occlusion-aware correction from a set of RS images. Bai et al.[54] and Lao and Ait-Aider[4] developed a scanline-homography and an RS-homography to remove RS effects, respectively. Recent studies have also demonstrated that a distortion-free GS image can be produced from an RS stereo rig with different RS directions[30] or from a generalized/standard RS stereo camera[40, 41].

Deep learning methods. Recently, the success of deep learning in high-level vision tasks has been gradually extended to the RS image correction task, where a deep neural network is trained end-to-end to warp the RS image to its perspective GS counterpart. This essentially becomes an image-to-image translation problem. Typically, these networks consist of two main components: a pixel-wise motion estimation module and a GS frame synthesis module, as illustrated in Fig. 6. The pixel-wise motion estimation module is dedicated to estimating the pixel-wise motion field, which is then used to warp the appearance information of adjacent frames to the target GS instance; the GS frame synthesis module aims to aggregate the context information from coarse to fine and eventually decode the desired GS frame.

For single-frame RS image correction, Rengarajan et al.[49] proposed the first convolutional neural network to mitigate RS artifacts by assuming a simple affine model with only 2 DoF: one is an in-plane rotation, and the other is a horizontal translation. On this basis, the RS regeneration scheme was exploited in [76]. Zhuang et al.[6] extended [49] to learn the underlying scene structure and camera motion from a single RS image, followed by a post-processing step to generate a geometrically consistent image. Mo et al.[57] predicted the camera pose of the anchor scanline and then used a cubic spline to approximate the arbitrary-scanline camera pose, avoiding the reliance on a specific motion model.

Given two-frame RS images as input, Liu et al.[72] pro-

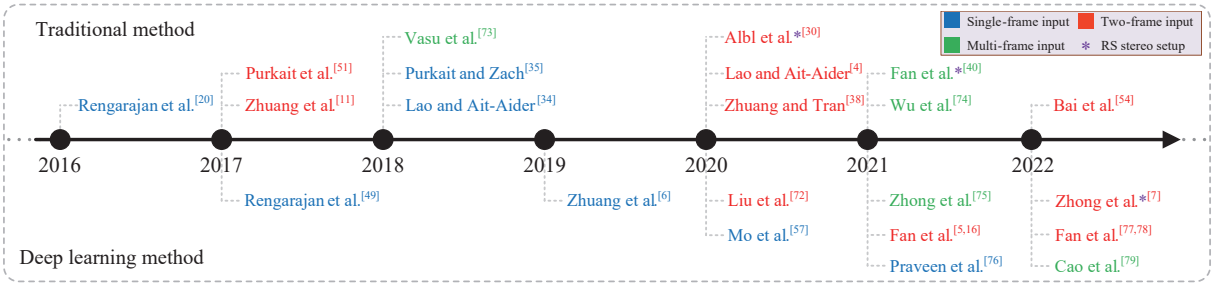


Fig. 5 Timeline of rolling shutter image correction methods

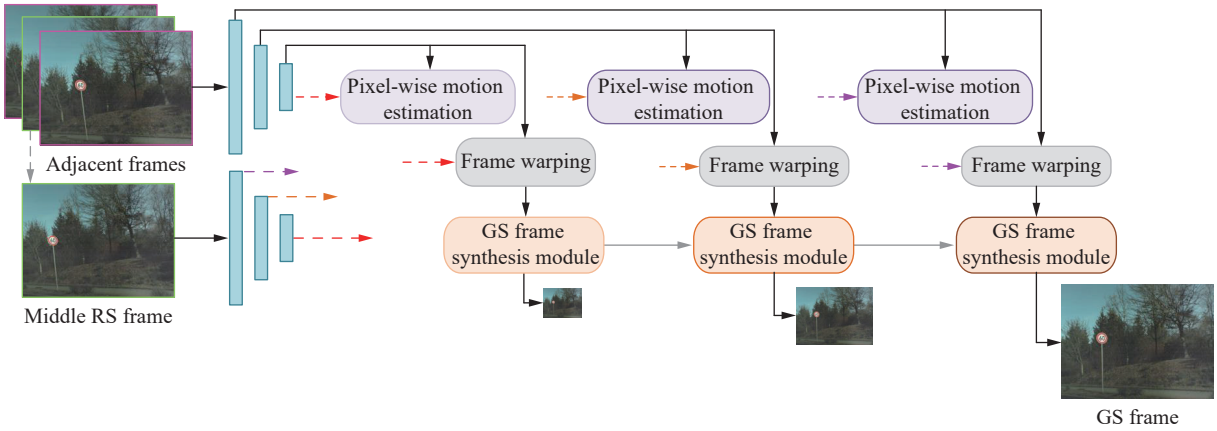


Fig. 6 Common framework for deep learning-based RS image correction methods

posed a deep shutter unrolling network (i.e., DeepUnrollNet) to recover the desired GS image from two consecutive RS images. Despite the promising performance, DeepUnrollNet solely uses the warped feature map corresponding to the second RS image when decoding the target GS frame, which tends to lead to content missing in the unseen regions of the recovered GS image. Very recently, to overcome this drawback, Fan et al.[16] designed a symmetric undistortion architecture by exploiting the temporal symmetry between two consecutive frames, where the contextual information is efficiently aggregated in a coarse-to-fine manner. Subsequently, the spatio-temporal consistency was essentially formulated by Fan et al.[5, 77, 78] to design a cascaded framework for simultaneous RS correction and temporal super-resolution.

With multi-frame RS images as input, Zhong et al.[75] implemented joint RS correction and image deblurring using three consecutive frames of the input, where a deformation attention module is designed to self-adaptively fuse deblurring and correction cues. Afterward, Zhong et al.[7] developed an end-to-end network for dual RS images with reversed RS directions, which is beneficial to remove the dependence of continuous RS frame input on the readout time ratio. Very recently, Cao et al.[79] presented an RS image correction method with adaptive warping to remove RS artifacts in real scenarios, where multiple displacement fields were estimated from three consecutive RS images for coarse-to-fine refinement.

Tables 4 and 5 uniformly summarize the quantitative performance of the classic methods for RS effect removal. The evaluation is conducted as follows: the Carla-RS dataset[72] with occlusion mask (CRM), the Carla-RS dataset without occlusion mask (CR), and the Fastec-RS dataset (FR)[72]. The average peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS)[87] are adopted for evaluation. A higher PSNR/SSIM or lower LPIPS score indicates better performance. We also report comparative results with the two-stage approaches. The two-stage methods indicate that the RS image correction methods (e.g., DeepUnrollNet[72], SUNet[16]) are performed first and then the video frame interpolation methods (e.g., BMBC[85], DAIN[86]) are adopted to generate the intermediate distortion-free frames. However, due to error accumulation, the two-stage methods tend to have blurring artifacts and local errors, as demonstrated in [5, 77, 78].

4.2 RS temporal super-resolution

A rolling shutter image may be viewed as a row-wise combination of a sequence of global shutter images captured by a (virtual) moving global shutter camera within the exposure duration. By exploiting the hidden-and-previously overlooked temporal-dynamic information, one could possibly invert the RS imaging mechanism to bring RS images alive. For example, a smooth high framerate GS video can be generated from two consecutive RS frames (figuratively, producing 1 440 GS video frames from two 720-height RS images), which is termed RS temporal super-resolution (RSSR), as shown in Fig. 7. In

Table 4 Quantitative comparisons on recovering GS images corresponding to the first scanline of the second RS frame (i.e., the exposure time $t = 0.5$). The four blocks from top to bottom are the traditional and learning-based RS correction methods, the two-stage method, and the RSSR method. We use gray backgrounds to mark the methods that can only produce one GS image at a specific exposure time (i.e., the RS image correction method). The average runtime for recovering a 640 × 480 GS image and the number of model parameters for each method are also reported. The runtime of the learning-based method is tested on an NVIDIA RTX 2 080 GPU.

Methods	Runtime (Second)	Params (Million)	PSNR ↑ (dB)			SSIM ↑		LPIPS ↓	
			CRM	CR	FR	CR	FR	CR	FR
DiffSfM[11]	467.3	–	24.20	21.28	20.14	0.775	0.701	0.132 2	0.178 9
DeepUnrollNet[72]	0.343	3.91	26.90	26.46	26.52	0.807	0.792	0.070 3	0.122 2
SUNet[16]	0.212	12.0	29.28	29.18	28.34	0.850	0.837	0.065 8	0.120 5
DeepUnrollNet[72] + BMBC[85]	2.250	14.9	27.29	27.58	24.95	0.829	0.787	0.098 0	0.202 4
DeepUnrollNet[72] + DAIN[86]	0.657	27.9	27.48	27.88	26.19	0.874	0.807	0.082 1	0.145 3
RSSR[5]	0.124	26.0	30.17	24.78	21.23	0.867	0.776	0.069 5	0.165 9
CVR[77]	0.141	42.6	32.02	31.74	28.72	0.929	0.847	0.036 8	0.110 7

Table 5 Quantitative comparisons on recovering GS images corresponding to the middle scanline of the second RS frame (i.e., the exposure time $t = 1.0$). The experiments are carried out on the Fastec-RS dataset[72]. The three blocks from top to bottom are the learning-based RS correction methods, the two-stage method, and the RSSR method.

Method	PSNR ↑ (dB)	SSIM ↑
DeepUnrollNet[72]	27.02	0.828
SUNet[16]	27.06	0.825
JCD[75]	26.48	0.821
AWNet[79]	28.56	0.855
SUNet[16] + BMBC[85]	25.49	0.796
SUNet[16] + DAIN[86]	27.12	0.823
RSSR[5]	24.89	0.824
CVR[77]	26.67	0.838

particular, for two rolling shutter frames $\boldsymbol{I}_0^r$ and $\boldsymbol{I}_1^r$ at adjacent times 0 and 1, the goal is to synthesize an intermediate GS frame $\boldsymbol{I}_t^g$, $t \in [0, 1]$.

As a joint interpolation and correction task, it is extremely challenging. One needs not only to eliminate geometric RS distortion but also to generate a set of high framerate GS images in chronological order. Meanwhile, it opens up opportunities for many practical applications such as computational photography, visual tracking, scene understanding, video editing and compression. Retracing the developments in recent years, flow-based deep learning solutions have received attention, as illustrated in Fig. 8. Note that the RS undistortion flow, including forward warping based $\boldsymbol{U}_{r\to g}$ from the RS image to the GS image and backward warping based $\boldsymbol{U}_{g\to r}$ from the GS image to the RS image, is defined and leveraged in these works. Fig. 9 shows the scanline-dependent properties of RS undistortion flows.

We report the quantitative results of GS video restoration for time steps $t = 0.5$ and $t = 1.0$ in Tables 4 and 5, respectively. Note that although [16, 72] can also recover GS images corresponding to different time steps, this requires training separate models based on the available supervised signals. Namely, they are far from the ability to generate smooth GS videos.

Fan and Dai[5] proposed the first learning-based RSSR solution for latent GS video extraction from two consecutive RS images. Under the assumption of a constant velocity of camera motion and a static scene, they demonstrated that interconversion between RS undistortion flows and regular optical flows, as well as between different RS undistortion flows corresponding to different target-scanlines, can be performed by scaling operations. Specifically, to transform each RS pixel x exposed at time τ_i, $i \in \{0, 1\}$ indicating the image index, to the GS canvas at time $t \in [0, 1]$, the forward warping based RS undistortion flow can be defined as

$$\begin{aligned} \boldsymbol{U}_{0\to t}(\boldsymbol{x}) &= \boldsymbol{C}_{0\to t}(\boldsymbol{x}) \odot \boldsymbol{F}_{0\to 1}(\boldsymbol{x}) \\ \boldsymbol{U}_{1\to t}(\boldsymbol{x}) &= \boldsymbol{C}_{1\to t}(\boldsymbol{x}) \odot \boldsymbol{F}_{1\to 0}(\boldsymbol{x}) \end{aligned} \tag{11}$$

which can be used directly to forward warp the RS image to the destination GS image. Here, $\odot$ indicates an element-wise multiplier, and

$$\begin{aligned} \boldsymbol{C}_{0\to t}(\boldsymbol{x}) &= \frac{(t-\tau_0)(h-\gamma\pi_v)}{h} \\ \boldsymbol{C}_{1\to t}(\boldsymbol{x}) &= \frac{(\tau_1-t)(h+\gamma\pi'_v)}{h} \end{aligned} \tag{12}$$

denote the correction maps. h represents the number of image scanlines. π_v and π'_v encapsulate the underlying RS geometry[5], which can be implicitly modeled by an encoder-decoder network (see Fig. 8(a)). Moreover, the conversion scheme between varying RS undistortion flows at times t_1 and t_2 can be obtained as

$$\boldsymbol{U}_{i\to t_2}(\boldsymbol{x}) = \frac{t_2-\tau}{t_1-\tau} \times \boldsymbol{U}_{i\to t_1}(\boldsymbol{x}), \quad i = 0, 1. \tag{13}$$

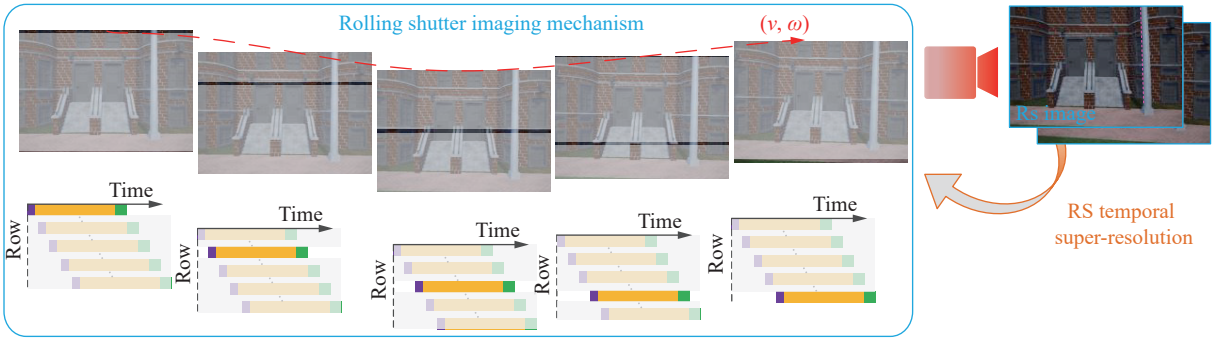


Fig. 7 Illustration of the RS temporal super-resolution task. This figure is from [5]. The RS image can be interpreted as the result of sequentially picking a row of pixels from the latent GS video frames during the exposure of the RS camera. By reversing the progressive imaging mechanism of the RS camera, RS temporal super-resolution can recover a smooth and coherent GS video sequence, making rolling shutter images vivid.

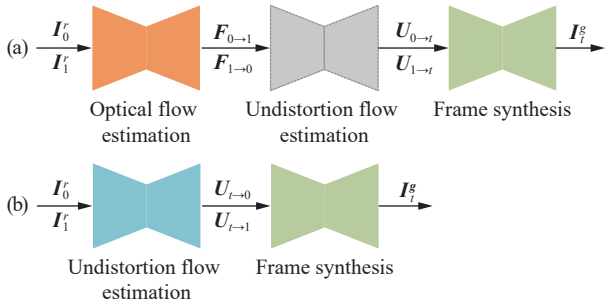


Fig. 8 Different flow-based RSSR paradigms. We roughly classify existing flow-based RSSR methods based on encoder-decoders with specific functions. In (a)[5, 77, 78], the optical flow network estimates regular optical flows $\boldsymbol{F}_{0\to1}$, $\boldsymbol{F}_{1\to0}$, the middle part generates RS undistortion flows $\boldsymbol{U}_{0\to t}$, $\boldsymbol{U}_{1\to t}$ via (11). Note that [5, 78] requires a specialized deep network (i.e., (12)), while [77] does not (i.e., (14)). In (b)[7], the backward warping-based RS undistortion flows $\boldsymbol{U}_{t\to0}$, $\boldsymbol{U}_{t\to1}$ are directly estimated by the network. A frame synthesis network is concatenated at the end of (a) and (b) for target GS image generation.

It is worth mentioning that the interconversion scheme under the constant acceleration motion assumption is extended in [78] and is not introduced here. Subsequently, Fan et al.[77] went further and proposed an approximate version of (12) by neglecting the parallax effects, i.e.,

$$\begin{aligned}\boldsymbol{C}_{0\to t}(\boldsymbol{x}) &= t-\tau_0\\ \boldsymbol{C}_{1\to t}(\boldsymbol{x}) &= \tau_1-t\end{aligned} \tag{14}$$

which are independent of image content and do not need to rely on specific neural networks. Meanwhile, a synthetic network is added at the end to aggregate contextual complementary information and improve robustness to occlusion and partially moving objects. Concurrently, as shown in Fig. 8(b), Zhong et al.[7] proposed directly learning the RS undistortion flow for backward warping, and then an encoder-decoder network was exploited to learn a synthetic mask for inferring the underlying GS image sequence. Note that instead of two consecutive RS frames as in [5, 77], Zhong et al.[7] explored a pair of images captured by dual RS cameras with reversed RS directions. The setup is the same as in [30].

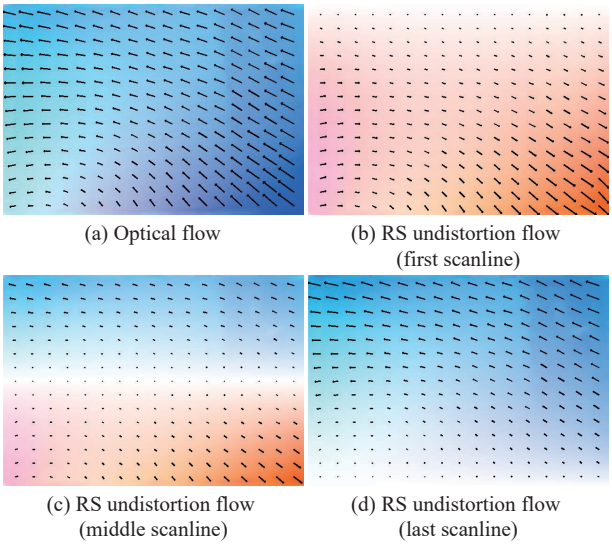


Fig. 9 Regular optical flow VS. RS undistortion flow. This figure is from [5], where the color indicates the direction of the flow, and the brightness indicates the magnitude of the flow. Unlike the isotropic smooth optical flow map, the RS undistorted flow map exhibits a more pronounced scanline dependence. The distribution of the magnitude and direction of the RS undistortion flow is related to the target scanline.

4.3 Public RS-based datasets

We classify publicly available RS-based datasets into three categories: RS correction datasets (which can be used for end-to-end training), cross-validation datasets (which can be used for performance evaluation), and datasets containing other degradations or data types (e.g., with motion blur, event/IMU data that can be used for multi-task processing).

4.3.1 RS correction datasets

Carla-RS.[72] It is generated from a virtual 3D environment using the Carla simulator[88], involving general 6-DoF camera motions. There are a training set of 210 se-

quences and a test set of 40 sequences, and each sequence consists of 10 consecutive frames. A total of 2 500 RS images with a resolution of 640×448 pixels are included.

Fastec-RS.[72] It uses a high-speed GS camera mounted on the ground vehicle to collect high-FPS GS video sequences at 2 400 Hz. Then, the RS image is synthesized by extracting pixels from consecutive GS images row-by-row and merging them. The training set has 56 sequences and the test set has 20 sequences, each of which contains 34 consecutive frames. There are 2 584 RS image pairs with a resolution of 640×480 pixels.

BS-RSC.[79] It is a realistic benchmark dataset, collected by a well-designed beam-splitter acquisition system in a dynamic urban environment. There are 50, 16, and 15 sequences for training (2 500 image pairs), validation (800 image pairs), and testing (750 image pairs), respectively. The image resolution is $1\,024 \times 768$ pixels.

4.3.2 Cross-validation datasets

Rolling shutter rectification dataset.[1][19] It has a small data scale with 6 evaluation sequences. Each sequence has 12 sets of data, each consisting of one RS frame and three ground truth GS frames. The camera motion contains pure rotation, pure translation, and varying readout time ratios (1.00, 0.96 and 0.92). Additionally, an RS video taken by an iPhone 3GS camera during fast motion is included. The image resolution is 640×480 pixels. This dataset can be used to evaluate the RS correction algorithm.

Rolling shutter bundle adjustment odometry dataset.[2][14] 36 video sequences are captured by an iPhone 4 camera at $1\,280 \times 720$ resolution. The frame rate is 30 Hz with a readout time of 32.37 microseconds. However, there is no corresponding accurate GS ground truth, so it can exclusively be used for qualitative assessment[11].

4.3.3 Datasets with other degradations or data types

BS-RSCD.[75] As a real dataset with ego-motion and object-motion, it is collected using a well-designed beam-splitter acquisition system. It can be used for simultaneous RS effect removal and deblurring tasks. The camera frame rate is 15 Hz. There are 50 sequences for training, 15 sequences for validation, and 15 sequences for testing. Each sequence has 50 video frames, i.e., 4 000 image pairs are recorded in total. The image resolution is 640×480 pixels.

RSGR-GS.[89] It is acquired using a well-designed beam-splitter acquisition system in real-world scenarios. Note that the RS camera employs a global reset mode for exposure. It contains 79 video sequences (27 for training, 3 for validation, and 49 for testing) with a resolution of 640×640 pixels.

Gev-RS.[90] This dataset records high-FPS GS video using a high-speed GS camera with a framerate of 5 700 Hz. Then, RS frames are synthesized similar to the Fastec-RS dataset and the corresponding event streams are simulated using V2E[91]. Finally, 3 700 "GS-event-RS" triplets containing 29 sequences are generated. The image resolution is 640×360 pixels.

TUM-RSVI.[10] As a rolling shutter visual-inertial odometry dataset[3], 10 sequences are collected by a multi-sensor setup, where time-synchronized RS and GS images (20 Hz), IMU data (200 Hz) and ground truth 6-DoF poses (120 Hz) are contained. The image resolution is $1\,280 \times 1\,024$ pixels. This dataset can be used for visual odometry, SLAM, etc.

5 Challenges and new trends

Despite the success of deep learning in RS correction and RS temporal super-resolution tasks, there are various open research questions regarding the network model design as discussed below:

1) Lighter and more efficient models. Existing network architectures stack a large number of 2D convolutional modules to essentially achieve image-to-image translation, and thus are not yet capable of real-time GS image recovery (cf. Table 4), especially on low-power mobile devices. In addition, limited by the low resolution of the current training dataset, it will be a challenge to design lighter network models for high-resolution images (e.g., 4 K video). As a result, designing more efficient network models to accelerate the inference will be crucial for real-time computer vision applications, such as visual SLAM[9, 92, 93].

2) Improve the generalization ability of the model. Since the RS image in the current dataset has a fixed readout time ratio, this may lead to poor generalization of the trained model to third-party RS cameras with significantly different readout time ratios. A straightforward approach is to enhance the diversity of the training data. However, there is little research on this topic and further research is needed.

3) Implement RS image correction together with other data/tasks. Currently, the performance of RS image correction is improved by combining it with event data[90], global reset[89], deblurring[75], etc. A future trend of data-driven models will be to associate other data types (e.g., IMU[94], depth camera[95], etc.) or other low-level image processing tasks (e.g., spatial super-resolution, spatio-temporal super-resolution, image denoising, radial distortion removal, etc.).

4) Generate more realistic and multi-instant training datasets. Training learning-based RS temporal super-resolution methods demands a significant number of RS and GS image pairs. The current datasets either use a beam-splitter acquisition system to obtain ground truth GS images of real scenes, e.g., [75, 79], or simulate RS images by stitching row-by-row with high framerate GS videos,

[1] https://www.cvl.isy.liu.se/research/datasets/rs-dataset

[2] https://www.cvl.isy.liu.se/research/datasets/rsba-dataset

[3] https://vision.in.tum.de/data/datasets/rolling-shutter-dataset

e.g., [7, 72]. However, the former only can capture one GS image corresponding to a single instant, which is severely insufficient for the RS temporal super-resolution task; the latter tends to produce striping artifacts as discussed in [75, 79], making the synthesized RS image insufficient to represent the real environment. To unleash the potential of deep learning methods, it is necessary to generate large-scale realistic RS training datasets with more exposure instants, more diverse scenes, and more dynamic objects.

6 Conclusions

This paper has provided a comprehensive survey of RS cameras. The geometric modeling and nonlinear optimization of RS motion have been provided mathematically, based on which three typical geometric applications have been pointed out to recover more accurate multi-view geometry. The taxonomy is carried out by considering RS model types and application types, where existing related work is further classified and presented. Furthermore, advances in deep learning techniques for RS image correction and RS temporal super-resolution are comprehensively summarized. This systematic taxonomy allows a better understanding of the principles and characteristics of RS models and provides a comprehensive guide for beginners to use RS cameras more efficiently. Finally, based on a review of the existing work, possible directions and open problems for RS cameras have also been discussed, with the aim of providing insight into the ongoing development of RS-related research. We hope that this survey will help systematize the existing work and spark a new wave of research in this long-standing field.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (Nos. 62271410, 61901387 and 62001394), the Fundamental Research Funds for the Central Universities, China, and the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University, China (No. CX2022046).

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge, UK: Cambridge University Press, 2003.

[2] Y. C. Dai, H. D. Li, L. Kneip. Rolling shutter camera relative pose: Generalized epipolar geometry. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 4132–4140, 2016. DOI: 10.1109/CVPR.2016.448.

[3] C. Albl, Z. Kukelova, T. Pajdla. R6P-rolling shutter absolute pose problem. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 2292–2300, 2015. DOI: 10.1109/CVPR.2015.7298842.

[4] Y. Z. Lao, O. Ait-Aider. Rolling shutter homography and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2780–2793, 2021. DOI: 10.1109/TPAMI.2020.2977644.

[5] B. Fan, Y. C. Dai. Inverting a rolling shutter camera: Bring rolling shutter images to high framerate global shutter video. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, IEEE, Montreal, Canada, pp. 4208–4217, 2021. DOI: 10.1109/ICCV48922.2021.00419.

[6] B. B. Zhuang, Q. H. Tran, P. Ji, L. F. Cheong, M. Chandraker. Learning structure-and-motion-aware rolling shutter correction. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Long Beach, USA, pp. 4546–4555, 2019. DOI: 10.1109/CVPR.2019.00468.

[7] Z. H. Zhong, M. D. Cao, X. Sun, Z. R. Wu, Z. Y. Zhou, Y. Q. Zheng, S. Lin, I. Sato. Bringing rolling shutter images alive with dual reversed distortion. In *Proceedings of the 17th European Conference on Computer Vision*, Springer, Tel Aviv, Israel, pp. 233–249, 2022. DOI: 10.1007/978-3-031-20071-7_14.

[8] Y. Z. Lao, O. Ait-Aider, A. Bartoli. Solving rolling shutter 3D vision problems using analogies with non-rigidity. *International Journal of Computer Vision*, vol. 129, no. 1, pp. 100–122, 2021. DOI: 10.1007/s11263-020-01368-1.

[9] J. H. Kim, C. Cadena, I. Reid. Direct semi-dense SLAM for rolling shutter cameras. In *Proceedings of IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, pp. 1308–1315, 2016. DOI: 10.1109/ICRA.2016.7487263.

[10] D. Schubert, N. Demmel, L. von Stumberg, V. Usenko, D. Cremers. Rolling-shutter modelling for direct visual-inertial odometry. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Macau, China, pp. 2462–2469, 2019. DOI: 10.1109/IROS40897.2019.8968539.

[11] B. B. Zhuang, L. F. Cheong, G. H. Lee. Rolling-shutter-aware differential SfM and image rectification. In *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 948–956, 2017. DOI: 10.1109/ICCV.2017.108.

[12] O. Saurer, M. Pollefeys, G. H. Lee. Sparse to dense 3D reconstruction from rolling shutter images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 3337–3345, 2016. DOI: 10.1109/CVPR.2016.363.

[13] S. Im, H. Ha, G. Choe, H. G. Jeon, K. Joo, I. S. Kweon. Accurate 3D reconstruction from small motion clip for rolling shutter cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 4, pp. 775–787, 2019. DOI: 10.1109/TPAMI.2018.2819679.

[14] J. Hedborg, P. E. Forssen, M. Felsberg, E. Ringaby. Rolling shutter bundle adjustment. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, pp. 1434–1441, 2012. DOI: 10.1109/CVPR.2012.6247831.

[15] M. Meingast, C. Geyer, S. Sastry. Geometric models of rolling-shutter cameras, [Online], Available: https://arxiv.org/abs/0503076, 2005.

[16] B. Fan, Y. C. Dai, M. Y. He. SUNet: Symmetric undistortion network for rolling shutter correction. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, IEEE, Montreal, Canada, pp. 4521–4530, 2021. DOI: 10.1109/ICCV48922.2021.00450.

[17] O. Saurer, K. Köser, J. Y. Bouguet, M. Pollefeys. Rolling shutter stereo. In *Proceedings of IEEE International Conference on Computer Vision*, Sydney, Australia, pp. 465–472, 2013. DOI: 10.1109/ICCV.2013.64.

[18] O. Saurer, M. Pollefeys, G. H. Lee. A minimal solution to the rolling shutter pose estimation problem. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Hamburg, Germany, pp. 1328–1334, 2015. DOI: 10.1109/IROS.2015.7353540.

[19] P. E. Forssén, E. Ringaby. Rectifying rolling shutter video from hand-held devices. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, pp. 507–514, 2010. DOI: 10.1109/CVPR.2010.5540173.

[20] V. Rengarajan, A. N. Rajagopalan, R. Aravind. From bows to arrows: Rolling shutter rectification of urban scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 2773–2781, 2016. DOI: 10.1109/CVPR.2016.303.

[21] L. Magerand, A. Bartoli, O. Ait-Aider, D. Pizarro. Global optimization of object pose and motion from a single rolling shutter image with automatic 2D-3D matching. In *Proceedings of the 12th European Conference on Computer Vision*, Springer, Florence, Italy, pp. 456–469, 2012. DOI: 10.1007/978-3-642-33718-5_33.

[22] K. Wang, B. Fan, Y. C. Dai. Relative pose estimation for stereo rolling shutter cameras. In *Proceedings of IEEE International Conference on Image Processing*, Abu Dhabi, UAE, pp. 463–467, 2020. DOI: 10.1109/ICIP40778.2020.9191254.

[23] O. Ait-Aider, N. Andreff, J. M. Lavest, P. Martinet. Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In *Proceedings of the 9th European Conference on Computer Vision*, Springer, Graz, Austria, pp. 56–68, 2006. DOI: 10.1007/11744047_5.

[24] O. Ait-Aider, F. Berry. Structure and kinematics triangulation with a rolling shutter stereo rig. In *Proceedings of the 12th International Conference on Computer Vision*, IEEE, Kyoto, Japan, pp. 1835–1840, 2009. DOI: 10.1109/ICCV.2009.5459408.

[25] C. Albl, Z. Kukelova, T. Pajdla. Rolling shutter absolute pose problem with known vertical direction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 3355–3363, 2016. DOI: 10.1109/CVPR.2016.365.

[26] C. Albl, A. Sugimoto, T. Pajdla. Degeneracies in rolling shutter SfM. In *Proceedings of the 14th European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, pp. 36–51, 2016. DOI: 10.1007/978-3-319-46454-1_3.

[27] Z. Kukelova, C. Albl, A. Sugimoto, T. Pajdla. Linear solution to the minimal absolute pose rolling shutter problem. In *Proceedings of the 14th Asian Conference on Computer Vision*, Springer, Perth, Australia, pp. 265–280, 2019. DOI: 10.1007/978-3-030-20893-6_17.

[28] C. Albl, Z. Kukelova, V. Larsson, T. Pajdla. Rolling shutter camera absolute pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, pp. 1439–1452, 2020. DOI: 10.1109/TPAMI.2019.2894395.

[29] Z. Kukelova, C. Albl, A. Sugimoto, K. Schindler, T. Pajdla. Minimal rolling shutter absolute pose with unknown focal length and radial distortion. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 698–714, 2020. DOI: 10.1007/978-3-030-58558-7_41.

[30] C. Albl, Z. Kukelova, V. Larsson, M. Polic, T. Pajdla, K. Schindler. From two rolling shutters to one global shutter. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Seattle, USA, pp. 2502–2510, 2020. DOI: 10.1109/CVPR42600.2020.00258.

[31] K. Wang, C. H. Liu, K. X. Wang, S. J. Shen. Depth estimation under motion with single pair rolling shutter stereo images. *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3160–3167, 2021. DOI: 10.1109/LRA.2021.3063695.

[32] E. Ringaby, P. E. Forssén. Efficient video rectification and stabilisation for cell-phones. *International Journal of Computer Vision*, vol. 96, no. 3, pp. 335–352, 2012. DOI: 10.1007/s11263-011-0465-8.

[33] E. Ito, T. Okatani. Self-calibration-based approach to critical motion sequences of rolling-shutter structure from motion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 4512–4520, 2017. DOI: 10.1109/CVPR.2017.480.

[34] Y. Z. Lao, O. Ait-Aider. A robust method for strong rolling shutter effects correction using lines with automatic feature selection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp. 4795–4803, 2018. DOI: 10.1109/CVPR.2018.00504.

[35] P. Purkait, C. Zach. Minimal solvers for monocular rolling shutter compensation under ackermann motion. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, Lake Tahoe, USA, pp. 903–911, 2018. DOI: 10.1109/WACV.2018.00104.

[36] C. R. Lee, J. H. Yoon, M. G. Park, K. J. Yoon. Gyroscope-aided relative pose estimation for rolling shutter cameras, [Online], Available: https://arxiv.org/abs/1904.06770, 2019.

[37] J. Hedborg, E. Ringaby, P. E. Forssén, M. Felsberg. Structure and motion estimation from rolling shutter video. In *Proceedings of IEEE International Conference on Computer Vision Workshops*, Barcelona, Spain, pp. 17–23, 2011. DOI: 10.1109/ICCVW.2011.6130217.

[38] B. B. Zhuang, Q. H. Tran. Image stitching and rectification for hand-held cameras. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 243–260, 2020. DOI: 10.1007/978-3-030-58571-6_15.

[39] S. Im, H. Ha, G. Choe, H. G. Jeon, K. Joo, I. S. Kweon. High quality structure from small motion for rolling shutter cameras. In *Proceedings of IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 837–845, 2015. DOI: 10.1109/ICCV.2015.102.

[40] B. Fan, Y. C. Dai, K. Wang. Rolling-shutter-stereo-aware motion estimation and image correction. *Computer Vision and Image Understanding*, vol. 213, Article number 103296, 2021. DOI: 10.1016/j.cviu.2021.103296.

[41] B. Fan, Y. C. Dai, Z. Y. Zhang, K. Wang. Differential SfM and image correction for a rolling shutter stereo rig. *Image*

and Vision Computing, vol. 124, Article number 104492, 2022. DOI: 10.1016/j.imavis.2022.104492.

[42] Y. Z. Lao, O. Ait-Aider, H. Araujo. Robustified structure from Motion with rolling-shutter camera using straightness constraint. *Pattern Recognition Letters*, vol. 111, pp. 1–8, 2018. DOI: 10.1016/j.patrec.2018.04.004.

[43] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon. Bundle adjustment − A modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms*, Springer, Corfu, Greece, pp. 298–372, 2000. DOI: 10.1007/3-540-44480-7_21.

[44] L. Oth, P. Furgale, L. Kneip, R. Siegwart. Rolling shutter camera calibration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Portland, USA, pp. 1360–1367, 2013. DOI: 10.1109/CVPR.2013.179.

[45] B. Y. Liao, D. L. Qu, Y. F. Xue, H. Q. Zhang, Y. Z. Lao. Revisiting rolling shutter bundle adjustment: Toward accurate and fast solution, [Online], Available: https://arxiv.org/abs/2209.08503, 2022.

[46] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011. DOI: 10.1007/s11263-010-0390-2.

[47] H. C. Longuet-Higgins, K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society B: Biological Sciences*, vol. 208, no. 1173, pp. 385–397, 1980. DOI: 10.1098/rspb.1980.0057.

[48] B. Fan, Y. C. Dai, Z. Y. Zhang, M. Y. He. Fast and robust differential relative pose estimation with radial distortion. *IEEE Signal Processing Letters*, vol. 29, pp. 294–298, 2022. DOI: 10.1109/LSP.2021.3134593.

[49] V. Rengarajan, Y. Balaji, A. N. Rajagopalan. Unrolling the shutter: CNN to correct motion distortions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 2345–2353, 2017. DOI: 10.1109/CVPR.2017.252.

[50] S. C. Su, W. Heidrich. Rolling shutter motion deblurring. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 1529–1537, 2015. DOI: 10.1109/CVPR.2015.7298760.

[51] P. Purkait, C. Zach, A. Leonardis. Rolling shutter correction in Manhattan world. In *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 882–890, 2017. DOI: 10.1109/ICCV.2017.101.

[52] A. Patron-Perez, S. Lovegrove, G. Sibley. A spline-based trajectory representation for sensor fusion and rolling shutter cameras. *International Journal of Computer Vision*, vol. 113, no. 3, pp. 208–219, 2015. DOI: 10.1007/s11263-015-0811-3.

[53] C. Kerl, J. Stückler, D. Cremers. Dense continuous-time tracking and mapping with rolling shutter RGB-D cameras. In *Proceedings of IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 2264–2272, 2015. DOI: 10.1109/ICCV.2015.261.

[54] F. Bai, A. Sengupta, A. Bartoli. Scanline homographies for rolling-shutter plane absolute pose. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, New Orleans, USA, pp. 8983–8992, 2022. DOI: 10.1109/CVPR52688.2022.00879.

[55] B. Vandeportaele, P. A. Gohard, M. Devy, B. Coudrin. Pose interpolation for rolling shutter cameras using non uniformly time-sampled B-splines. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, SciTePress, Porto, Portugal, pp. 286–293, 2017. DOI: 10.5220/0006171802860293.

[56] S. Lovegrove, A. Patron-Perez, G. Sibley. Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras. In *Proceedings of the British Machine Vision Conference*, BMVA Press, Bristol, UK, pp. 93.1–93.12, 2013. DOI: 10.5244/C.27.93.

[57] J. Mo, J. Islam, J. Sattar. Learning rolling shutter correction from real data without camera motion assumption, [Online], Available: https://arxiv.org/abs/2011.03106, 2020.

[58] J. Z. Huai, Y. K. Zhuang, Y. Lin, G. Jozkow, Q. C. Yuan, D. Chen. Continuous-time spatiotemporal calibration of a rolling shutter camera-IMU system. *IEEE Sensors Journal*, vol. 22, no. 8, pp. 7920–7930, 2022. DOI: 10.1109/JSEN.2022.3152572.

[59] X. L. Lang, J. J. Lv, J. X. Huang, Y. K. Ma, Y. Liu, X. X. Zuo. Ctrl-VIO: Continuous-time visual-inertial odometry for rolling shutter cameras. *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11537–11544, 2022. DOI: 10.1109/LRA.2022.3202349.

[60] E. B. Dam, M. Koch, M. Lillholm. Quaternions, Interpolation and Animation, Technical Report DIKU-TR-98/5, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark, 1998.

[61] C. H. Zhao, B. Fan, J. W. Hu, Q. Pan, Z. Xu. Homography-based camera pose estimation with known gravity direction for UAV navigation. *Science China Information Sciences*, vol. 64, no. 1, Article number 112204, 2021. DOI: 10.1007/s11432-019-2690-0.

[62] C. H. Zhao, B. Fan, J. W. Hu, L. M. Tian, Z. Y. Zhang, S. J. Li, Q. Pan. Pose estimation for multi-camera systems. In *Proceedings of IEEE International Conference on Unmanned Systems*, Beijing, China, pp. 533–538, 2017. DOI: 10.1109/ICUS.2017.8278403.

[63] S. C. Zhou, R. Yan, J. X. Li, Y. K. Chen, H. J. Tang. A brain-inspired SLAM system based on orb features. *International Journal of Automation and Computing*, vol. 14, no. 5, pp. 564–575, 2017. DOI: 10.1007/s11633-017-1090-y.

[64] C. L. Wang, T. M. Wang, J. H. Liang, Y. C. Zhang, Y. Zhou. Bearing-only visual SLAM for small unmanned aerial vehicles in GPS-denied environments. *International Journal of Automation and Computing*, vol. 10, no. 5, pp. 387–396, 2013. DOI: 10.1007/s11633-013-0735-8.

[65] Y. Z. Lao, O. Ait-Aider, A. Bartoli. Rolling shutter pose and ego-motion estimation using shape-from-template. In *Proceedings of the 15th European Conference on Computer Vision*, Springer, Munich, Germany, pp. 477–492, 2018. DOI: 10.1007/978-3-030-01216-8_29.

[66] O. Ait-Aider, A. Bartoli, N. Andreff. Kinematics from lines in a single rolling shutter image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007. DOI: 10.1109/CVPR.2007.383119.

[67] R. M. Haralick, D. Lee, K. Ottenburg, M. Nolle. Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Maui, USA, pp. 592–598, 1991. DOI: 10.1109/CVPR.1991.139759.

[68] M. Panda, B. Das, B. Subudhi, B. B. Pati. A comprehensive review of path planning algorithms for autonomous underwater vehicles. *International Journal of Automation and Computing*, vol. 17, no. 3, pp. 321–352, 2020. DOI: 10.1007/s11633-019-1204-9.

[69] Y. Yang, F. Qiu, H. Li, L. Zhang, M. L. Wang, M. Y. Fu. Large-scale 3D semantic mapping using stereo vision. *International Journal of Automation and Computing*, vol. 15, no. 2, pp. 194–206, 2018. DOI: 10.1007/s11633-018-1118-y.

[70] Q. Qi, Q. D. Li, Y. Q. Cheng, Q. Q. Hong. Skeleton marching-based parallel vascular geometry reconstruction using implicit functions. *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 30–43, 2020. DOI: 10.1007/s11633-019-1189-4.

[71] B. Fan, K. Wang, Y. C. Dai, M. Y. He. RS-DPSNet: Deep plane sweep network for rolling shutter stereo images. *IEEE Signal Processing Letters*, vol. 28, pp. 1550–1554, 2021. DOI: 10.1109/LSP.2021.3099350.

[72] P. D. Liu, Z. P. Cui, V. Larsson, M. Pollefeys. Deep shutter unrolling network. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Seattle, USA, pp. 5940–5948, 2020. DOI: 10.1109/CVPR42600.2020.00598.

[73] S. Vasu, M. R. M. Mohan, A. N. Rajagopalan. Occlusion-aware rolling shutter rectification of 3D scenes. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp. 636–645, 2018. DOI: 10.1109/CVPR.2018.00073.

[74] H. C. Wu, L. Xiao, Z. H. Wei. Simultaneous video stabilization and rolling shutter removal. *IEEE Transactions on Image Processing*, vol. 30, pp. 4637–4652, 2021. DOI: 10.1109/TIP.2021.3073865.

[75] Z. H. Zhong, Y. Q. Zheng, I. Sato. Towards rolling shutter correction and deblurring in dynamic scenes. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Nashville, USA, pp. 9215–9224, 2021. DOI: 10.1109/CVPR46437.2021.00910.

[76] K. Praveen, T Lokesh Kumar, A. N. Rajagopalan. Deep network for rolling shutter rectification, [Online], Available: https://arxiv.org/abs/2112.06170, 2021.

[77] B. Fan, Y. C. Dai, Z. Y. Zhang, Q. Liu, M. Y. He. Context-aware video reconstruction for rolling shutter cameras. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, New Orleans, USA, pp. 17551–17561, 2022. DOI: 10.1109/CVPR52688.2022.01705.

[78] B. Fan, Y. C. Dai, H. D. Li. Rolling shutter inversion: Bring rolling shutter images to high framerate global shutter video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to be published. DOI: 10.1109/TPAMI.2022.3212912.

[79] M. D. Cao, Z. H. Zhong, J. H. Wang, Y. Q. Zheng, Y. J. Yang. Learning adaptive warping for real world rolling shutter correction. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, New Orleans, USA, pp. 17764–17772, 2022. DOI: 10.1109/CVPR52688.2022.01726.

[80] C. K. Liang, L. W. Chang, H. H. Chen. Analysis and compensation of rolling shutter effect. *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1323–1330, 2008. DOI: 10.1109/TIP.2008.925384.

[81] S. Baker, E. Bennett, S. B. Kang, R. Szeliski. Removing rolling shutter wobble. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, USA, pp. 2392–2399, 2010. DOI: 10.1109/CVPR.2010.5539932.

[82] M. Grundmann, V. Kwatra, D. Castro, I. Essa. Calibration-free rolling shutter removal. In *Proceedings of IEEE International Conference on Computational Photography*, Seattle, USA, pp. 1–8, 2012. DOI: 10.1109/ICCPhot.2012.6215213.

[83] A. Punnappurath, V. Rengarajan, A. N. Rajagopalan. Rolling shutter super-resolution. In *Proceedings of IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 558–566, 2015. DOI: 10.1109/ICCV.2015.71.

[84] M. Meilland, T. Drummond, A. I. Comport. A unified rolling shutter and motion blur model for 3D visual registration. In *Proceedings of IEEE International Conference on Computer Vision*, Sydney, Australia, pp. 2016–2023, 2013. DOI: 10.1109/ICCV.2013.252.

[85] J. Park, K. Ko, C. Lee, C. S. Kim. BMBC: Bilateral motion estimation with bilateral cost volume for video interpolation. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 109–125, 2020. DOI: 10.1007/978-3-030-58568-6_7.

[86] W. B. Bao, W. S. Lai, C. Ma, X. Y. Zhang, Z. Y. Gao, M. H. Yang. Depth-aware video frame interpolation. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Long Beach, USA, pp. 3698–3707, 2019. DOI: 10.1109/CVPR.2019.00382.

[87] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp. 586–595, 2018. DOI: 10.1109/CVPR.2018.00068.

[88] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. López, V. Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, Mountain View, USA, vol. 78, pp. 1–16, 2017.

[89] Z. X. Wang, X. Ji, J. B. Huang, S. Satoh, X. Zhou, Y. Q. Zheng. Neural global shutter: Learn to restore video from a rolling shutter camera with global reset feature. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, New Orleans, USA, pp. 17773–17782, 2022. DOI: 10.1109/CVPR52688.2022.01727.

[90] X. Y. Zhou, P. Q. Duan, Y. Ma, B. X. Shi. EvUnroll: Neuromorphic events based rolling shutter image correction. In *Proceedings of IEEE/CVF conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 17754–17763, 2022. DOI: 10.1109/CVPR52688.2022.01725.

[91] Y. H. Hu, S. C. Liu, T. Delbruck. V2E: From video frames to realistic DVS events. In *Proceedings of IEEE/CVF conference on Computer Vision and Pattern Recognition Workshops*, IEEE, Nashville, USA, pp. 1312–1321, 2021. DOI: 10.1109/CVPRW53098.2021.00144.

[92] J. H. Kim, Y. Latif, I. Reid. RRD-SLAM: Radial-distorted rolling-shutter direct SLAM. In *Proceedings of IEEE International Conference on Robotics and Automation*, Singapore, pp. 5148–5154, 2017. DOI: 10.1109/ICRA.2017.7989602.

[93] D. Schubert, N. Demmel, V. Usenko, J. Stückler, D. Cremers. Direct sparse odometry with rolling shutter. In *Proceedings of the 15th European Conference on Computer*

Vision, Springer, Munich, Germany, pp. 699–714, 2018. DOI: 10.1007/978-3-030-01237-3_42.

[94] J. Mo, J. Islam, J. Sattar. IMU-assisted learning of single-view rolling shutter correction. In *Proceedings of the Conference on Robot Learning*, London, UK, vol. 164, pp. 861–870, 2021.

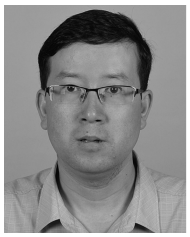
[95] S. Tourani, S. Mittal, A. Nagariya, V. Chari, M. Krishna. Rolling shutter and motion blur removal for depth cameras. In *Proceedings of IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, pp. 5098–5105, 2016. DOI: 10.1109/ICRA.2016.7487715.

Bin Fan received the B. Sc. degree in statistics and the M. Eng. degree in control science and engineering from Northwestern Polytechnical University, China in 2016 and 2019, respectively. He is currently a Ph. D. degree candiclate in information and communication engineering with School of Electronics and Information, Northwestern Polytechnical University (NPU), China. He was selected to the CVPR 2022 Doctoral Consortium (the only one among Chinese universities). He co-organized the ACCV 2022 tutorial on the topic of rolling shutter cameras. He has published some papers in TPAMI, CVPR, ICCV, TCSVT, CVIU, IVC, etc.

His research interests include computer vision, image processing, 3D reconstruction, and deep learning, especially regarding the rolling shutter camera.

E-mail: binfan@mail.nwpu.edu.cn

ORCID iD: 0000-0002-8028-0166



Yuchao Dai received the B. Eng., M. Eeg. and Ph. D. degrees all in signal and information processing from Northwestern Polytechnical University, China in 2005, 2008 and 2012, respectively. He is currently a professor with School of Electronics and Information, Northwestern Polytechnical University (NPU), China. He was an ARC DECRA fellow with the research school of engineering at Australian National University, Australia. He won the Best Paper Award in IEEE CVPR 2012, the Best Paper Award Nominee at IEEE CVPR 2020, the DSTO Best Fundamental Contribution to Image Processing Paper Prize at DICTA 2014, the Best Algorithm Prize in NRSFM Challenge at CVPR 2017, the Best Student Paper Prize at DICTA 2017, the Best Deep/Machine Learning Paper Prize at APSIPA ASC 2017. He served as Area Chair in CVPR, ICCV, ACM MM, ACCV, etc. He serves as Publicity Chair in ACCV 2022.

His research interests include structure from motion, multi-view geometry, low-level computer vision, deep learning, compressive sensing, and optimization.

E-mail: daiyuchao@nwpu.edu.cn (Corresponding author)

ORCID iD: 0000-0002-4432-7406

Mingyi He received the B. Eng. degree in electronic engineering and the M. Eng. degree in signal and systems from Northwestern Polytechnical University (NPU), China in 1982 and 1985, respectively, and the Ph. D. degree in signal and information processing from Xidian University, China in 1994. Since 1985, he has been with School of Electronics and Information, NPU, where he has been a full professor since 1996 and appointed as a chief professor of SIP in 1998. He was the (co)recipient of the 2012 CVPR Best Paper Award, the 2017 APSIPA ASC Best Deep/Machine Learning Paper Award, and the 2017 DICTA Best Student Paper Award. He was a recipient of the Government Lifelong Subsidy from the State Council of China and the Baosteel Outstanding Teacher Award in 2017. He received awards from the IEEE Signal Processing Society in 2014, APSIPA in 2019, China Remote Sensing Committee in 2023, *Journal of Image and Graphs* in 2022, *Signal Processing* in 2023, the Chinese Institute of Electronics in 2018 and 2020, and the Shaanxi Institute of Electronics in 2020. He has acted as the general chair or the TPC (co)chair and the area chair for over 30 national and international conferences. He was also an Associate Editor of the *IEEE Transactions on Geoscience and Remote Sensing* and APSIPA SIP and a Guest Editor of the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. He is Fellow of CIE and Vice President of APSIPA (2021-2024).

His research interests focus on advanced machine vision and intelligent processing, including signal and image processing, computer vision, hyper-spectral remote sensing, 3D information acquisition and processing, and neural network artificial intelligence.

E-mail: myhe@nwpu.edu.cn

ORCID iD: 0000-0003-2051-6955