西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

ACCV 2022 Macau

Rolling Shutter Camera: Modeling, Optimization and Learning




Yuchao Dai, Bin Fan

Northwestern Polytechnical University

daiyuchao@nwpu.edu.cn

Dec. 5, 2022

The Computer Vision and Robotics Group, CVR

Outline

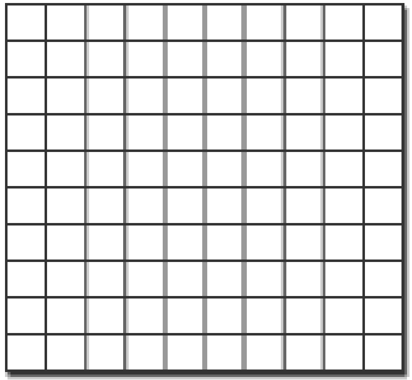
- **Introduction** (09:00-09:30)
- **Rolling Shutter Geometric Modeling and Optimization** (09:30-10:30)
  - Global Shutter Geometric Model
  - Rolling Shutter Uniform Motion Model
  - Rolling Shutter Differential Motion Model
  - Typical Applications
- **Learning-based Rolling Shutter Image Processing** (11:00-12:00)
  - Rolling Shutter Correction
  - Rolling Shutter Temporal Super-Resolution
  - Public Datasets
- **Further Direction and Discussion** (12:00-13:00)

1. Introduction

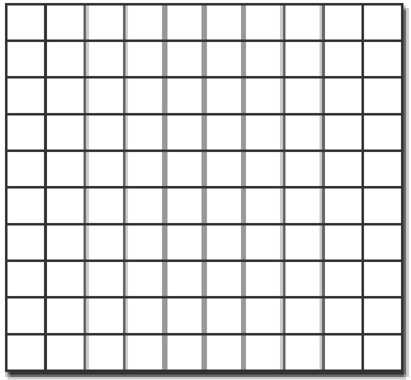
Rolling Shutter Effect

- Create some unintended geometric distortions if you’re filming fast-moving subjects or panning your video camera across a scene, such as skew, wobble, etc.
- Common in footage from DSLRs and mobile phone cameras.

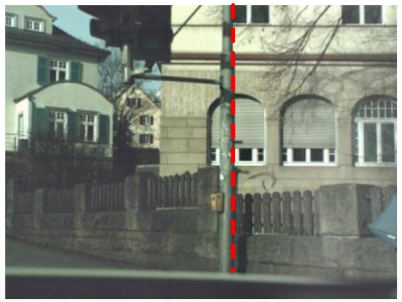
Rolling Shutter



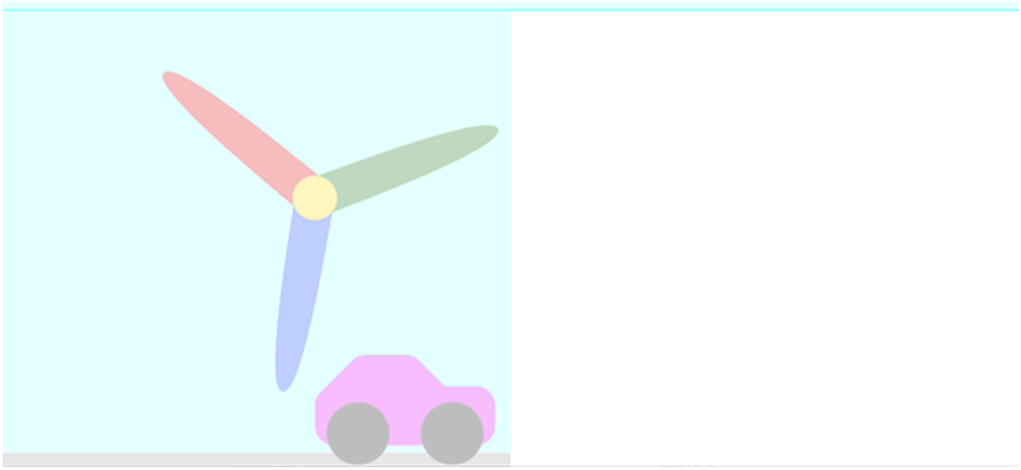
Global Shutter



Rolling shutter image



Global shutter image

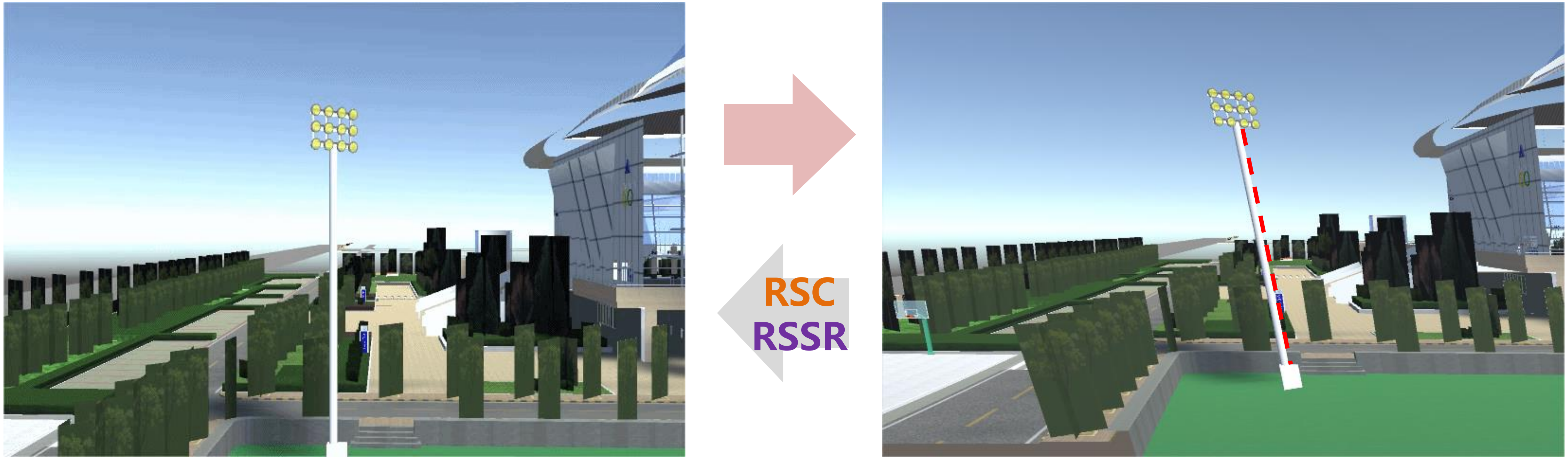


When RS effect relevant for computer vision:

- 3D modeling from images;
- Visual SLAM;
- Video stabilization, Video panorama, etc.;
- Any geometric measurement from images.

1. Introduction

Problem Formulation



Latent global shutter image sequence

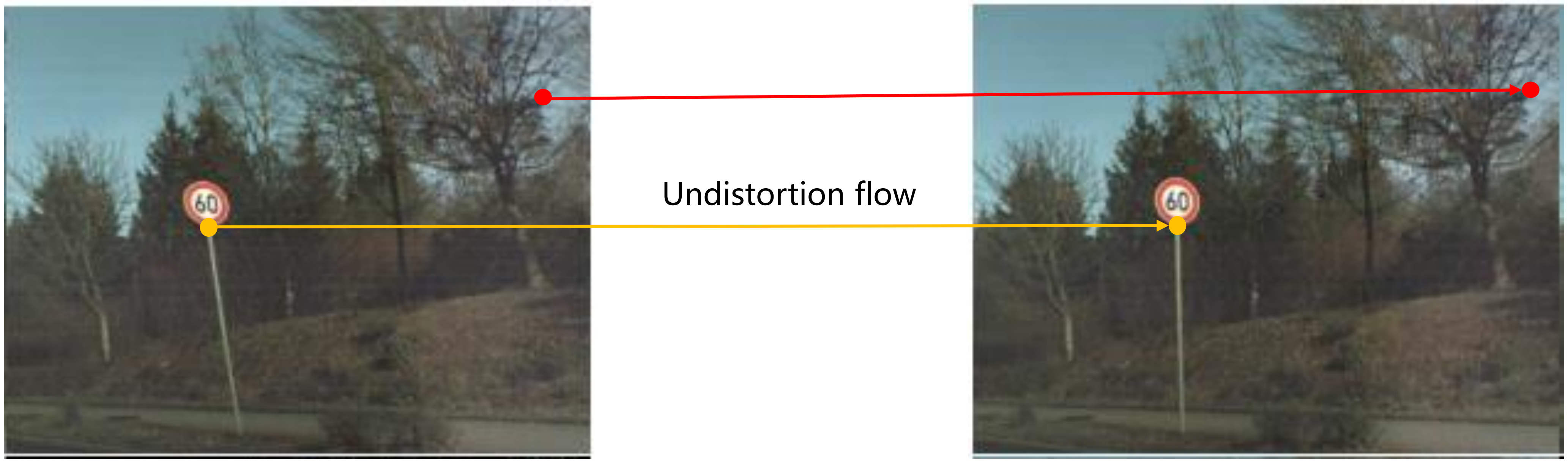
Rolling shutter image

- ✓ Rolling shutter images can be viewed as the result of the row-wise combination of global shutter images captured by a virtual moving GS camera during imaging.
- ✓ Rolling shutter images implicitly contain **rich high framerate temporal dynamic observation information**, i.e., camera motion information (temporally) and scene 3D information (spatially).
- ✓ Under the framework of temporal dynamic modeling and deep learning, recovering the global shutter image **corresponding to a specific exposure moment (i.e., Rolling Shutter Correction, RSC)** or **corresponding to any exposure moment (i.e., Rolling Shutter Temporal Super-Resolution, RSSR)** has become a research hotspot.

2. Rolling Shutter Correction

Formulation:

- Given a single or multiple rolling shutter images, we aim at estimating the undistortion flow to **recover a latent global shutter image corresponding to a specific exposure moment**, such as the first/middle scanline of the rolling shutter frame.



Rolling shutter (RS) image

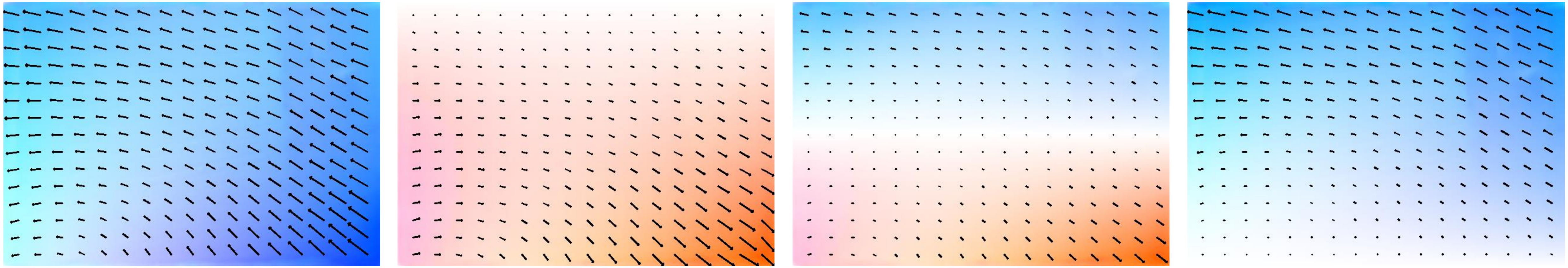
Global shutter (GS) image

2. Rolling Shutter Correction

Undistortion Flow vs. Optical Flow

- The undistortion flow map exhibits the significant **scanline dependence**.
- The undistortion flow near the target scanline appears as smaller warping displacement values;
- The undistortion flow corresponding to pixels that are opposite to the target scanline shows different warping displacement directions.

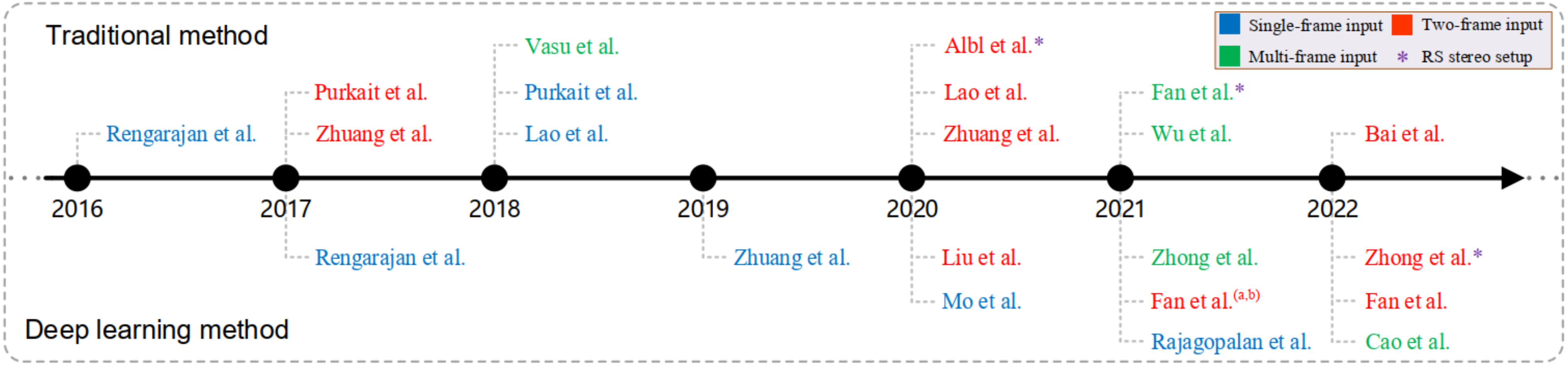
Optical Flow | Undistorted Flow (first scanline) | Undistorted Flow (middle scanline) | Undistorted Flow (last scanline)



2. Rolling Shutter Correction

Related work:

- Over the last decade, several **traditional** works usually rely on hand-designed prior assumptions, geometric constraints, and complex optimization frameworks to remove the rolling shutter effect.
- In recent years, several appealing **deep learning-based** rolling shutter correction methods have been proposed, where a convolutional neural network is trained to warp the rolling shutter frame to its global shutter counterpart. This essentially becomes an **image-to-image translation problem**.

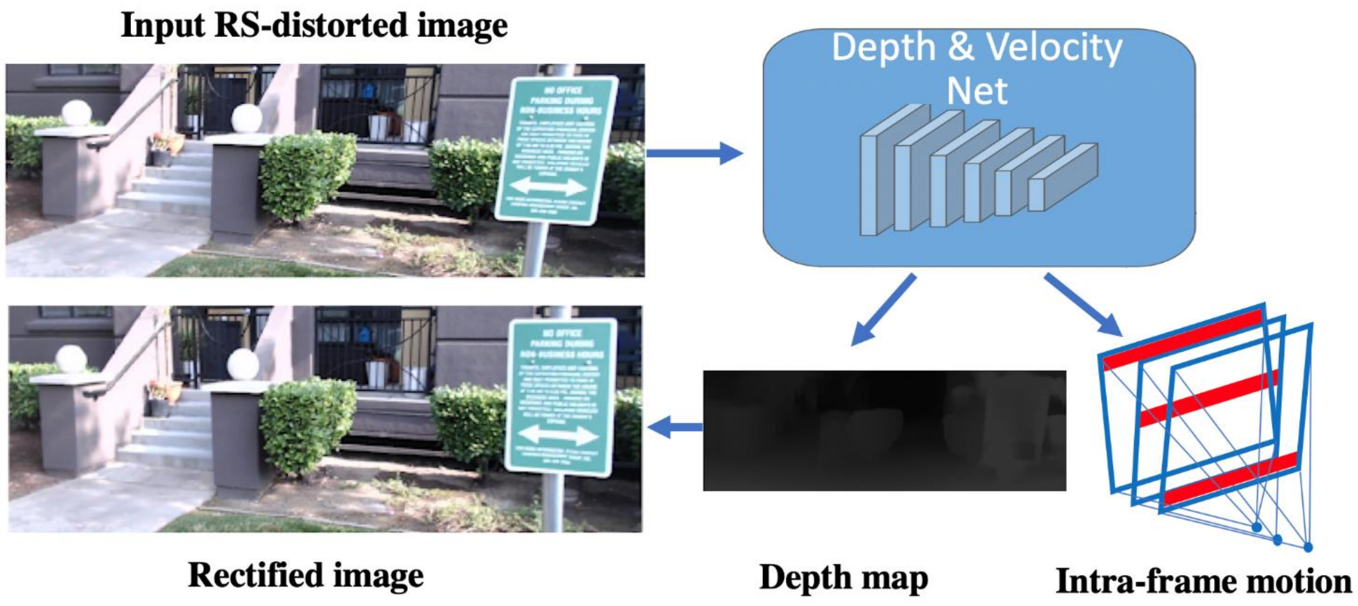


Timeline of rolling shutter image correction methods.

2. Rolling Shutter Correction

Related work:

- **Single-frame** rolling shutter correction is inherently **a highly ill-posed problem**.
- E.g., Zhuang et al. use data-driven priors through a network that learns the underlying scene depth and intra-frame motion from a single rolling shutter image, followed by a post-processing step to generate a geometrically consistent image.
- **Using at least two consecutive frames can make it tractable.**

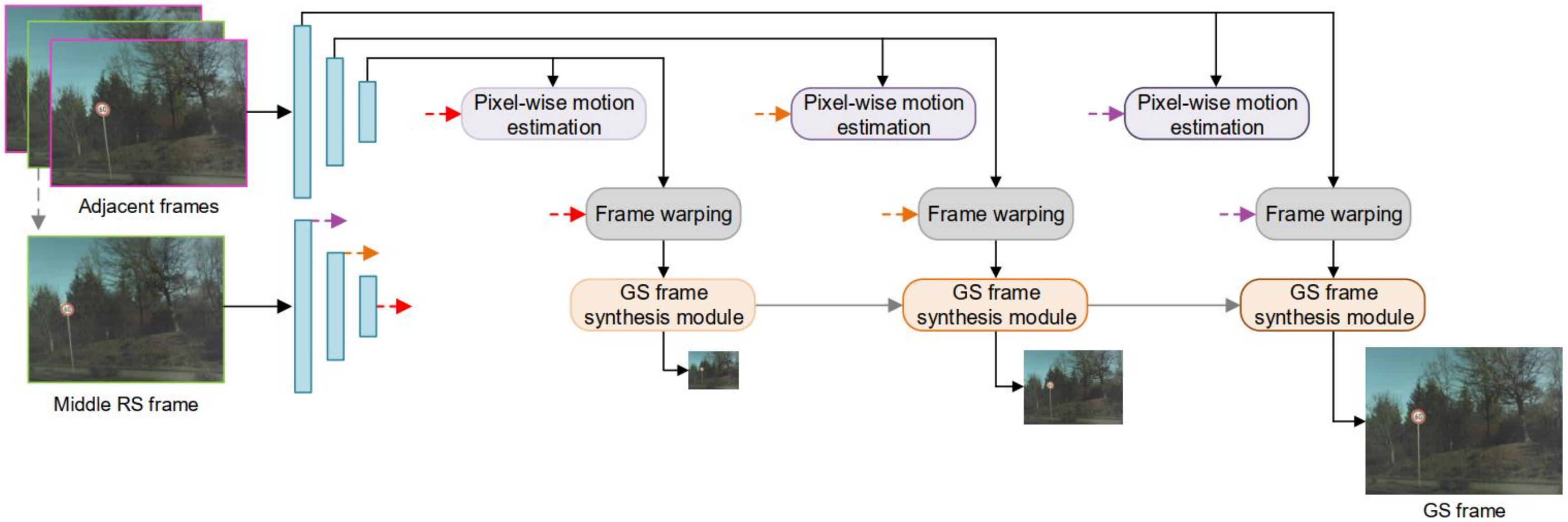


Zhuang B, Tran Q H, Ji P, et al. Learning structure-and-motion-aware rolling shutter. CVPR 2019.

2. Rolling Shutter Correction

Related work:

- These multi-frame-based methods consist of two main components: a **pixel-wise motion estimation module** and a **global shutter frame synthesis module**.
- The pixel-wise motion estimation module is dedicated to estimating the pixel-wise motion field, which is then used to warp the image appearance information of adjacent frames to the target global shutter instance;
- The global shutter frame synthesis module aims to aggregate the context information from coarse to fine and finally decode the desired global shutter image.



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

Deep Shutter Unrolling Network

Peidong Liu, Zhaopeng Cui, Viktor Larsson, Marc Pollefeys

CVPR 2020

2.1. Deep Shutter Unrolling Network

Pipeline:

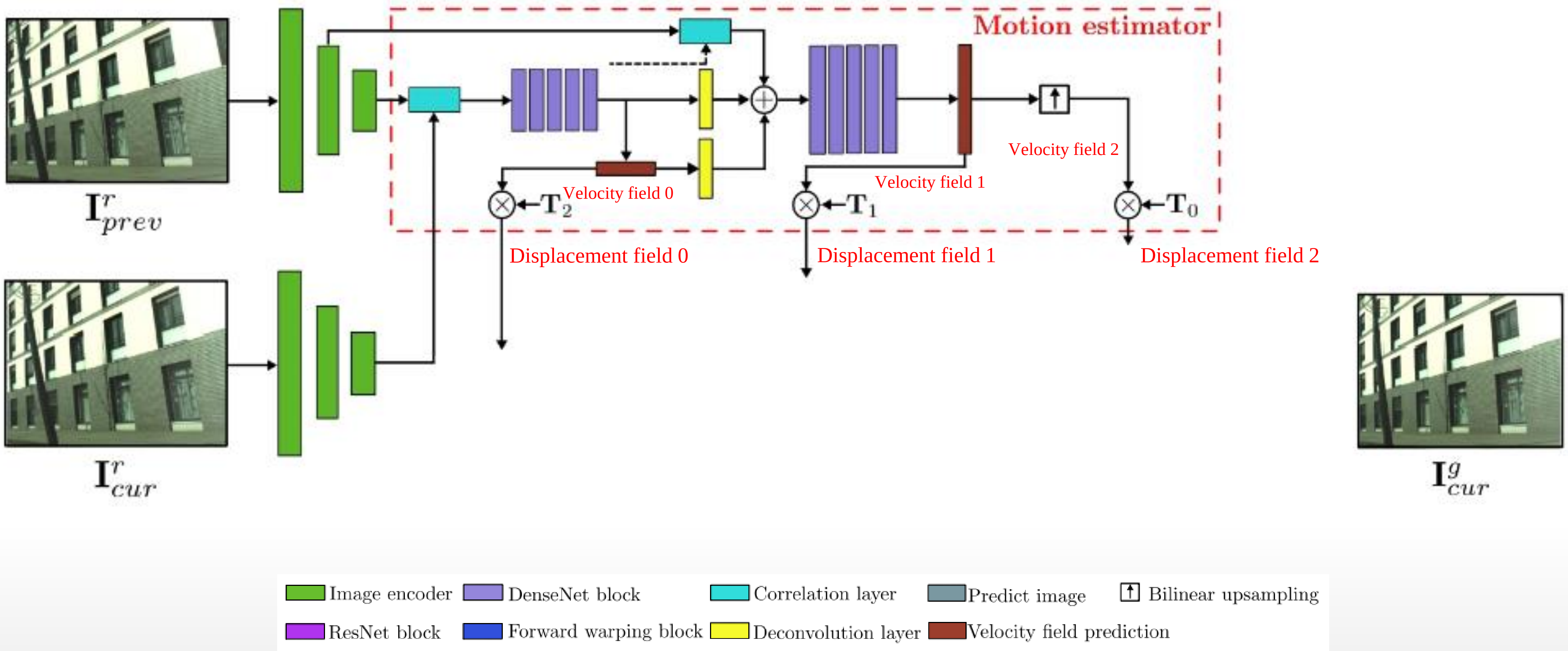
- Given two-frame rolling shutter images as input, Liu et al. proposed a deep shutter unrolling network (i.e. DeepUnrollNet) to recover the desired global shutter image from two consecutive rolling shutter images.

Liu P, Cui Z, Larsson V, et al. Deep shutter unrolling network. CVPR 2020.

2.1. Deep Shutter Unrolling Network

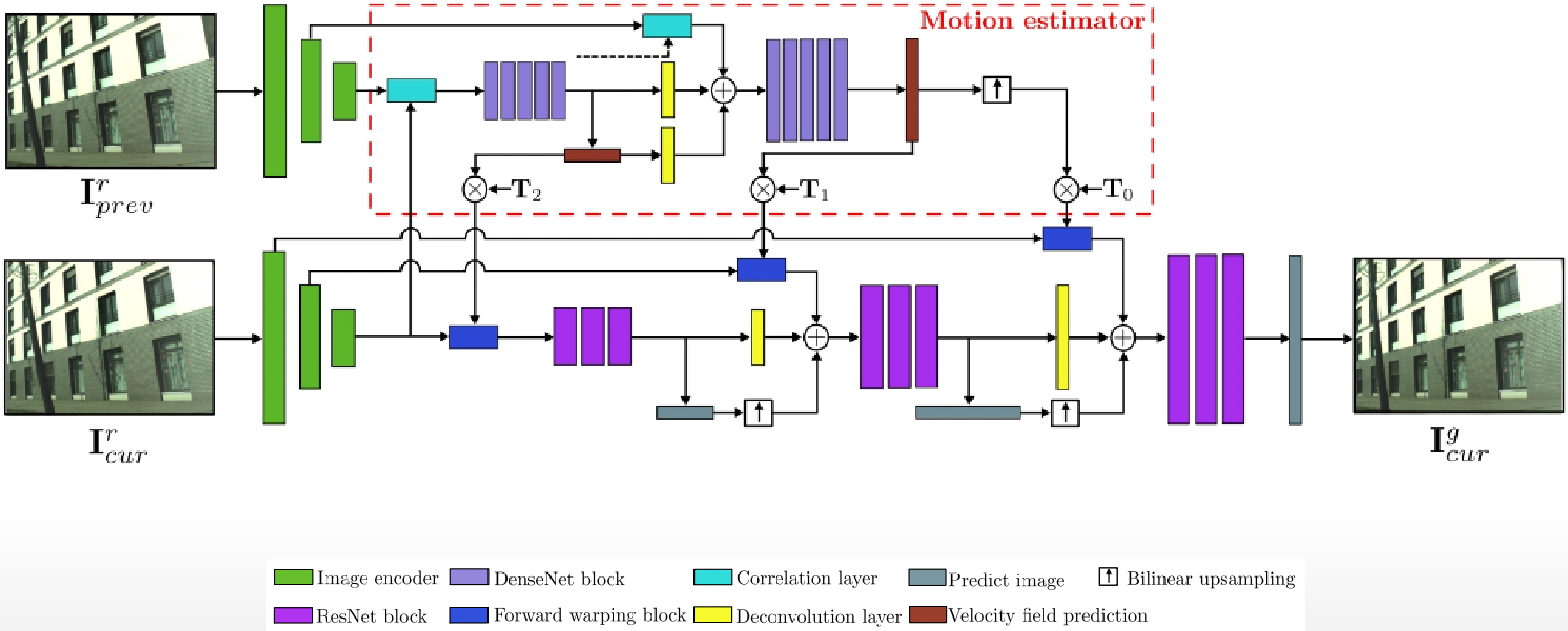
Pipeline:

- First, the information of the two frames is combined to estimate the undistortion flow, which can be inspired from the optical flow estimation network.



2.1. Deep Shutter Unrolling Network

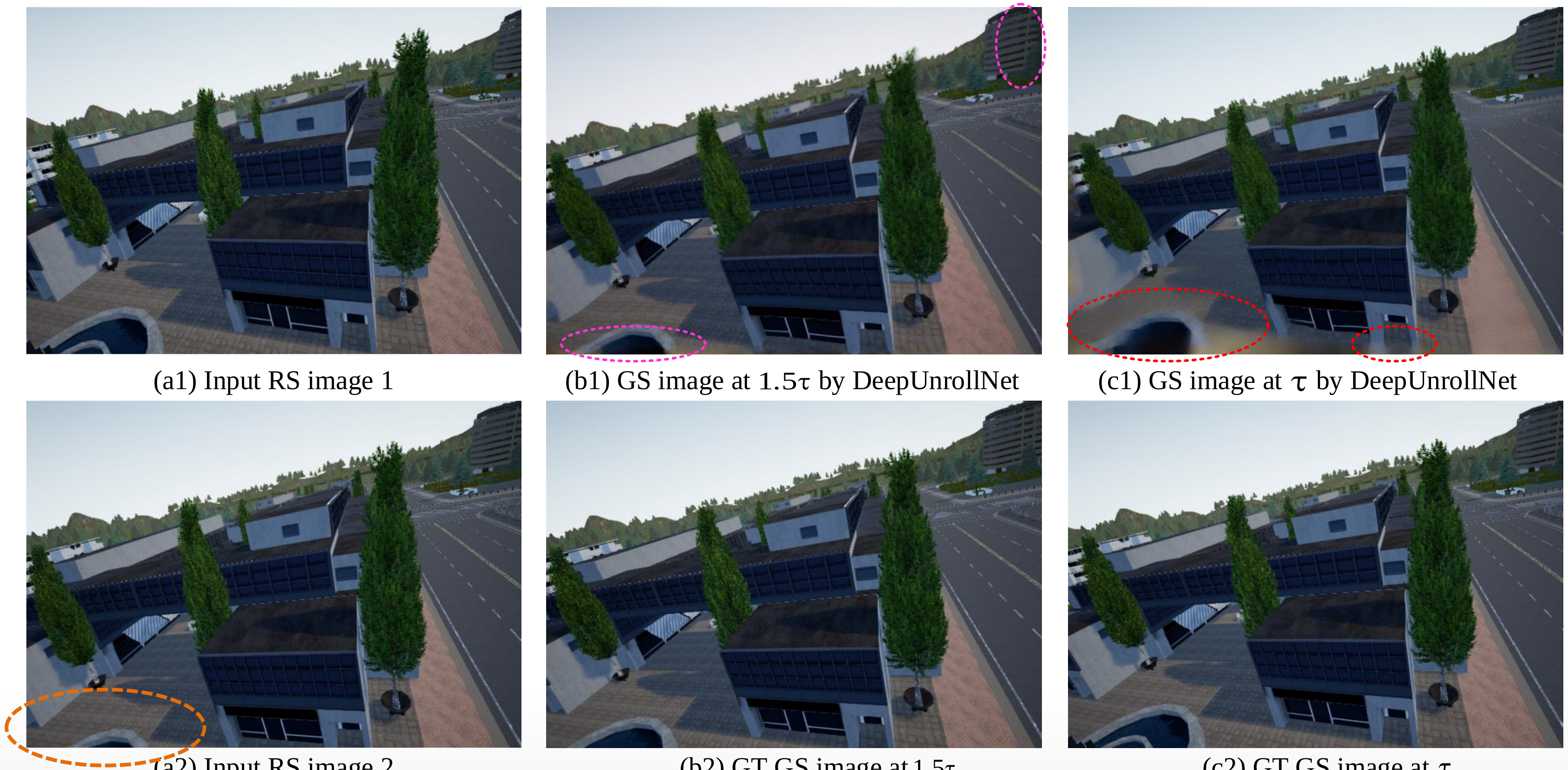
- **Pipeline:**
- Then, the image appearance features are warped to a common global shutter canvas, which is subsequently decoded to the target global shutter image in a coarse-to-fine manner.



2.1. Deep Shutter Unrolling Network

Shortcoming of DeepUnrollNet:

- Despite the promising performance, DeepUnrollNet solely uses the warped feature map corresponding to the second rolling shutter image when decoding the target global shutter frame, which tends to lead to content missing in the unseen regions of the recovered global shutter image.



(a1) Input RS image 1 | (b1) GS image at 1.5τ by DeepUnrollNet | (c1) GS image at τ by DeepUnrollNet

(a2) Input RS image 2 | (b2) GT GS image at 1.5τ | (c2) GT GS image at τ

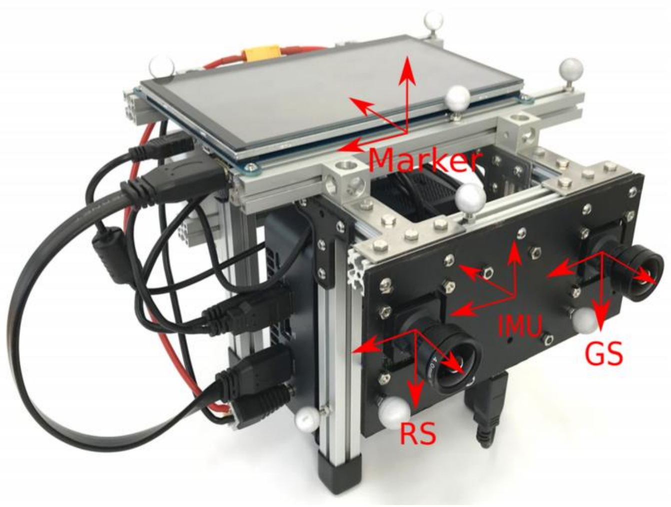
SUNet: Symmetric Undistortion Network for Rolling Shutter Correction

Bin Fan, Yuchao Dai, Mingyi He*

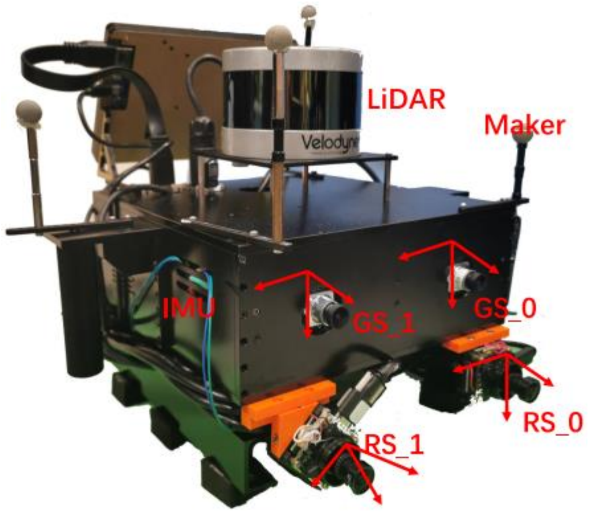
ICCV 2021

2.2. SUNet: Symmetric Undistortion Network for RS Correction

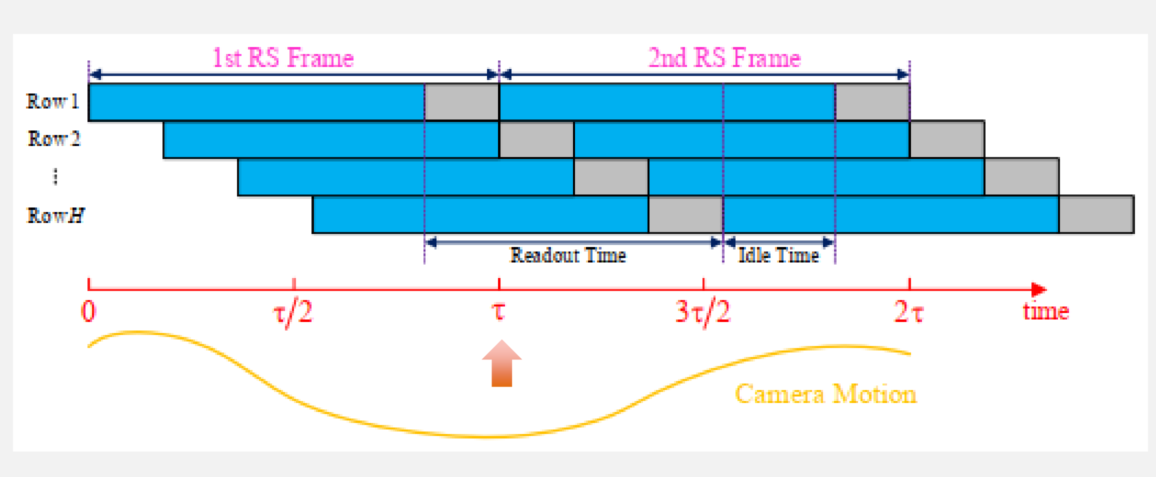
- **Background:**
  - RS cameras are usually time-synchronized with other sensors (e.g., GS camera, IMU, etc.) in hardware by referring to the first scanline time.
  - It is crucial and valuable to recover the GS image corresponding the first scanline of the second frame (i.e., the intermediate time **τ** of these two frames).



[Schubert et al, IROS’19]

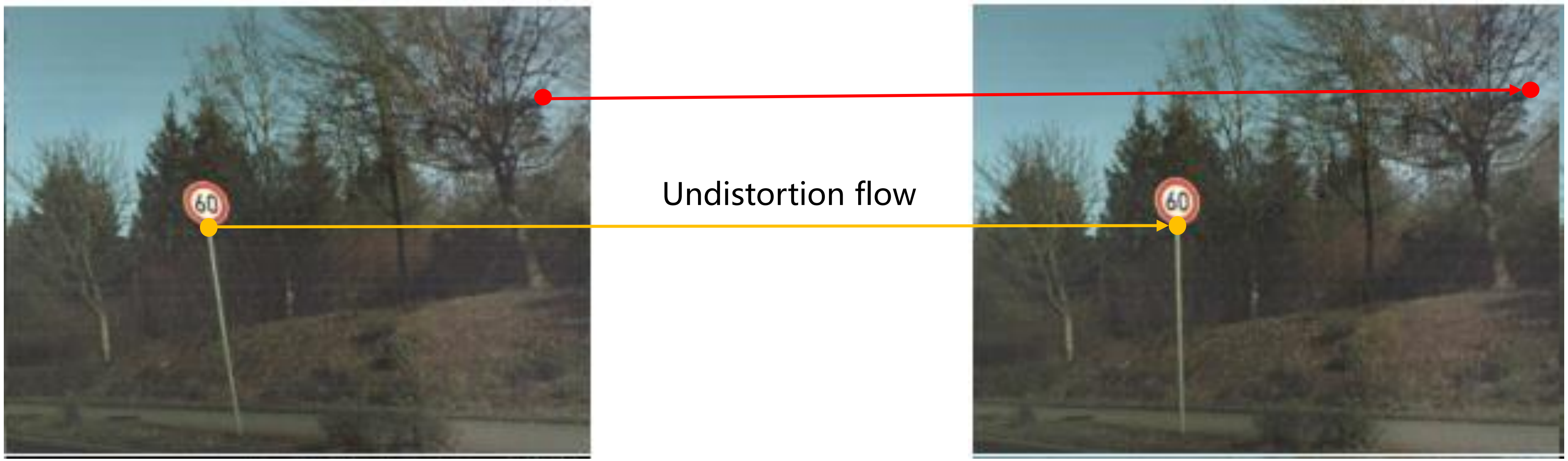


[Wang et al, RAL&ICRA’21]



2.2. SUNet: Symmetric Undistortion Network for RS Correction

- **Objective:**
  - Given *two* consecutive rolling shutter images, recover the global shutter image corresponding to the camera pose of **the first scanline of the second frame**.



Rolling shutter (RS) image

Global shutter (GS) image

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Challenges:

- **Maybe large pixel displacement** (e.g., foreground objects): The pixel of the target GS image may not be in the neighboring pixel of its corresponding RS image, depending on the type of motion, the 3D structure, and the scanline time.
- Due to the temporal continuity, we observe that the first and second RS images contribute greatly to the lower and upper parts of the corresponding time-centered GS image, respectively.

Original RS image 1 | Original RS image 2 | Ground truth GS image

Predicted only by RS 1 | Predicted only by RS 2 | Our corrected GS image

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Contributions:

- We propose an efficient end-to-end symmetric rolling shutter undistortion network to solve the generic RS correction problem with two consecutive frames.
- Our context-aware cost volume together with the symmetric consistency constraint can aggregate the contextual cues of two input RS images effectively.
- Our method significantly outperforms the state-of-the-art methods in both GS image restoration and inference efficiency.

Inputs

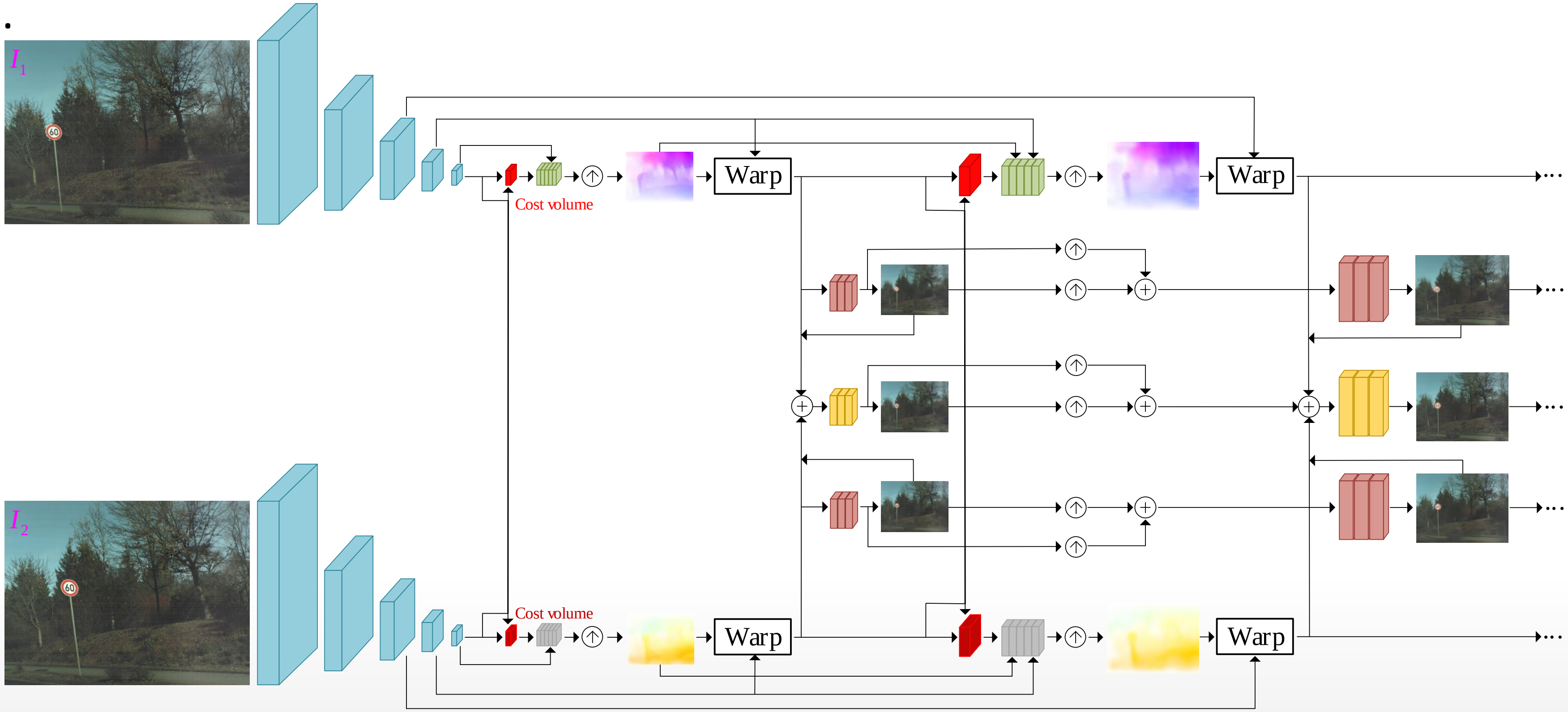
SUNet (Ours)

DeepUnrollNet

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Pipeline of our method:

- First, two time-symmetric dense **undistortion flows** are estimated by using well-established principles: pyramidal construction, warping, and cost volume processing.
- Then, both rolling shutter images are **warped** into a common global shutter one in the feature space.
- Finally, a symmetric consistency constraint is constructed in the **image decoder** to effectively aggregate the contextual cues of two RS images, thereby recovering the high-quality global shutter image.



2.2. SUNet: Symmetric Undistortion Network for RS Correction

Training loss: $\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_p \mathcal{L}_p + \lambda_c \mathcal{L}_c + \lambda_s \mathcal{L}_s$

- **Reconstruction loss**: evaluating the pixel-wise reconstruction quality of the corrected GS image on multiple scales

$$\mathcal{L}_r = \sum_{l=l_0-1}^{L} \left\| \boldsymbol{I}_{GT}^{l-1} - \boldsymbol{I}_g^{l-1} \right\|_1$$

- **Perceptual loss**: preserving details of the predictions and make estimated GS image sharper

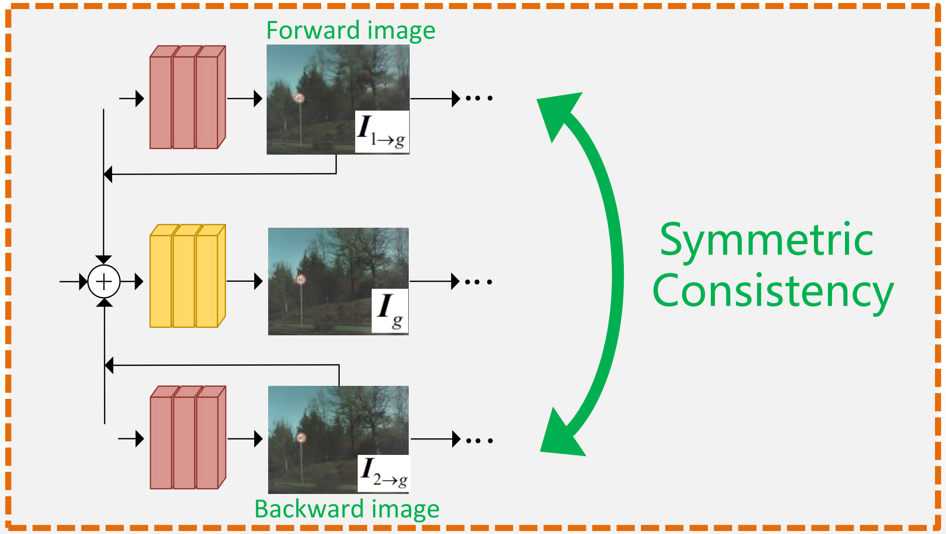
$$\mathcal{L}_p = \sum_{l=l_0-1}^{L} \left\| \phi\left(\boldsymbol{I}_{GT}^{l-1}\right) - \phi\left(\boldsymbol{I}_g^{l-1}\right) \right\|_1$$

- **Smoothness loss**: encouraging piecewise smoothness in the estimated undistortion flows

$$\mathcal{L}_s = \sum_{t=1}^{2} \sum_{l=l_0}^{L} \left\| \nabla \boldsymbol{F}_{t\to g}^{l-1} \right\|_2$$

- **Consistency loss**: To combine cues from two consecutive RS frames, we enforce their respective warped features to be as close to each other as possible in the symmetric space. I.e., we supervise the network to align the forward and backward images predicted by the first and the second RS images respectively across different levels

$$\mathcal{L}_c = \sum_{t=1}^{2} \sum_{l=l_0}^{L} \left\| \boldsymbol{I}_{GT}^{l-1} - \boldsymbol{I}_{t\to g}^{l-1} \right\|_1$$



2.2. SUNet: Symmetric Undistortion Network for RS Correction

Experiments

- Results on Carla-RS and Fastec-RS benchmarks

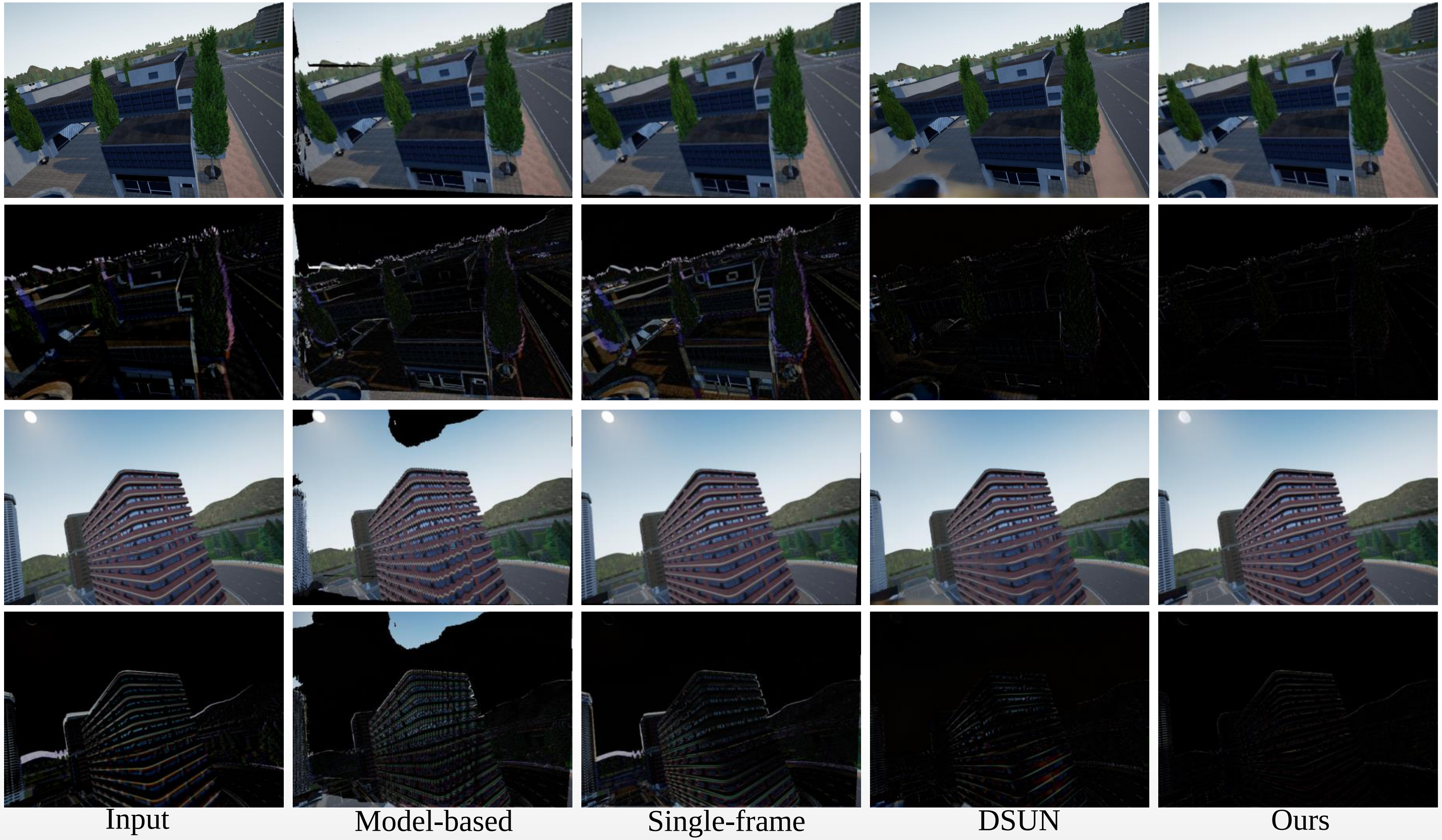
Methods	PSNR↑ (dB)			SSIM↑	
	CRM	CR	FR	CR	FR
Single-frame [34]	18.70	18.47	-	0.58	-
Model-based [32]	25.93	22.88	21.44	0.77	0.71
DSUN [18]	26.90	26.46	26.52	0.81	0.79
SUNet (Ours)	**29.28**	**29.18**	**28.34**	**0.85**	**0.84**

(a) Original RS image 2 (b) Ground truth GS image (c) Zhuang *et al.* [32] (d) Liu *et al.* [18] (e) Ours

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Experiments

- Qualitative comparison



2.2. SUNet: Symmetric Undistortion Network for RS Correction

Experiments

- Intermediate outputs of our method

(a) RS image 1: $\boldsymbol{I}_1$

(b) RS image 2: $\boldsymbol{I}_2$

(c) Forward undistortion flow: $\boldsymbol{F}_{1\rightarrow g}$

(d) Backward undistortion flow: $\boldsymbol{F}_{2\rightarrow g}$

(e) Forward GS image: $\boldsymbol{I}_{1\rightarrow g}$

(f) Backward GS image: $\boldsymbol{I}_{2\rightarrow g}$

(g) Our corrected GS image: $\boldsymbol{I}_g$

(h) Ground truth GS image: $\boldsymbol{I}_{GT}$

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Experiments

- Ablation study on loss function

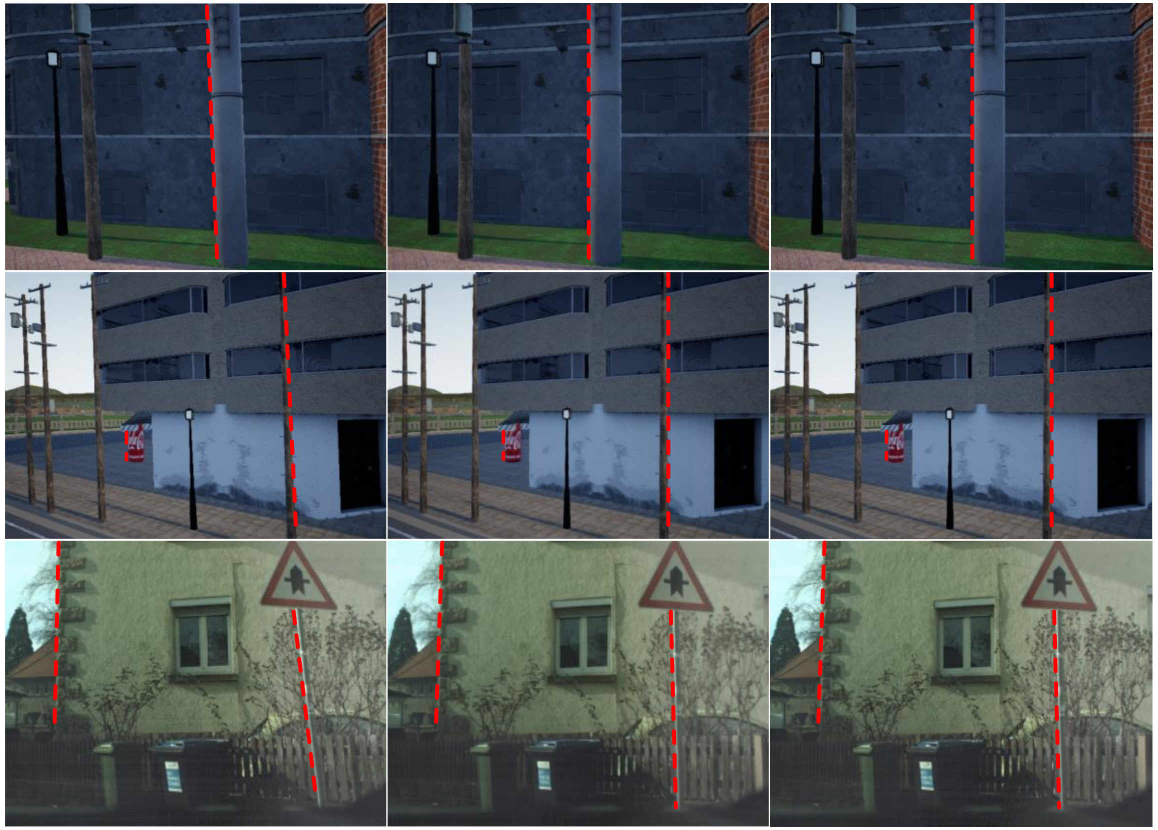
Table 2. Effectiveness of different combinations of training losses.

	PSNR↑			SSIM↑	
	CRM	CR	FR	CR	FR
w/o $\mathcal{L}_r$	28.00	27.90	27.29	0.83	0.81
w/o $\mathcal{L}_p$	29.08	28.95	28.20	**0.85**	**0.84**
w/o $\mathcal{L}_c$	29.05	28.94	27.89	0.84	0.82
w/o $\mathcal{L}_s$	29.19	28.07	28.15	**0.85**	0.83
full loss	**29.28**	**29.18**	**28.34**	**0.85**	**0.84**

Table 4. Ablation study on the consistency loss. $\lambda_c = 0$ means no consistency loss is used. The self-supervised consistency loss is defined as measuring only the difference between forward and backward GS images. Our loss function is effective to align contextual cues, especially the Fastec-RS dataset.

Consist. Loss	PSNR↑			SSIM↑	
	CRM	CR	FR	CR	FR
$\lambda_c = 0$	29.05	28.94	27.89	0.84	0.82
Self-supervised	29.15	28.99	28.02	**0.85**	0.83
Ours	**29.28**	**29.18**	**28.34**	**0.85**	**0.84**

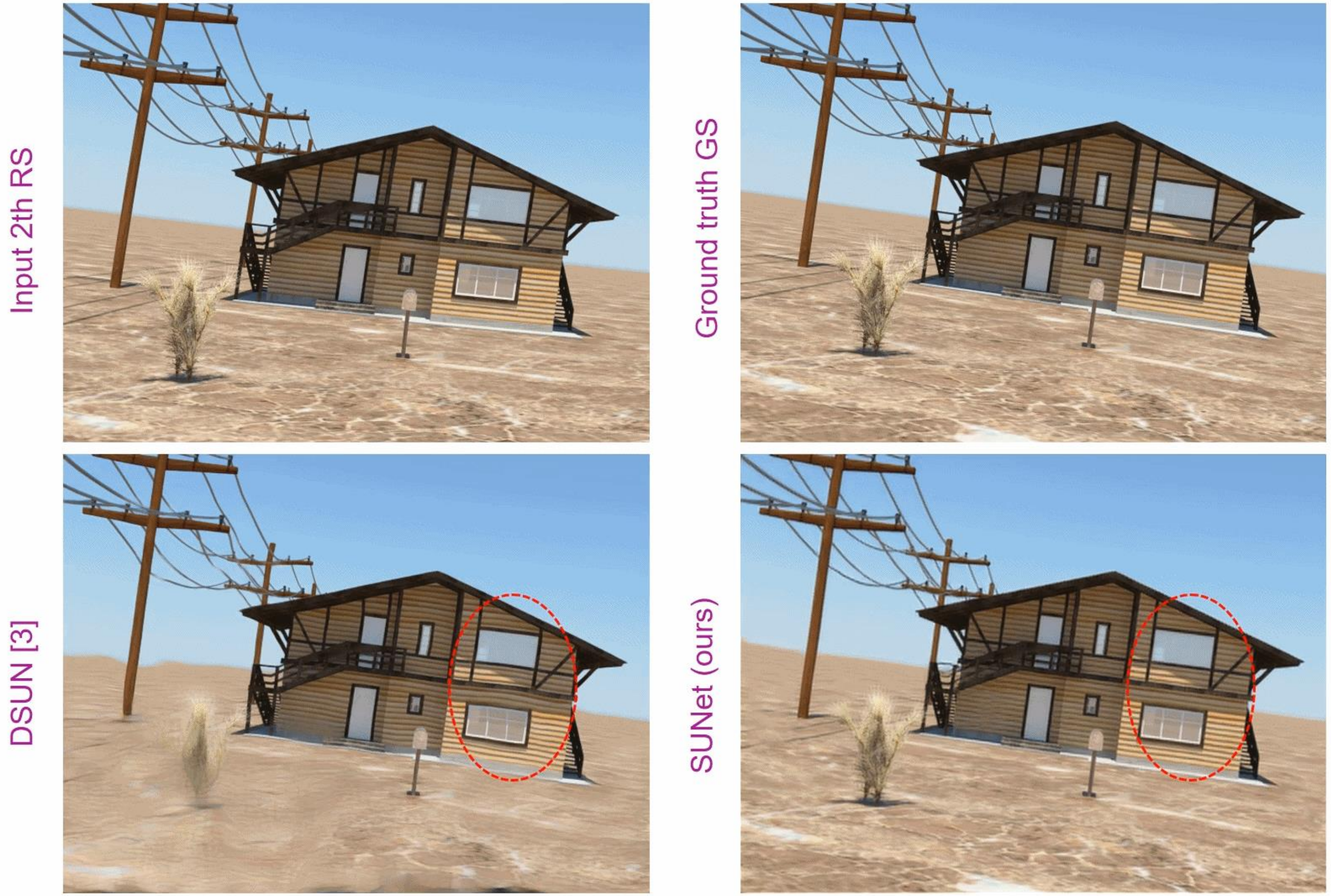
- Handle significant depth-dependent occlusion



2.2. SUNet: Symmetric Undistortion Network for RS Correctionv

Experiments

- Correction results on RS video[1]

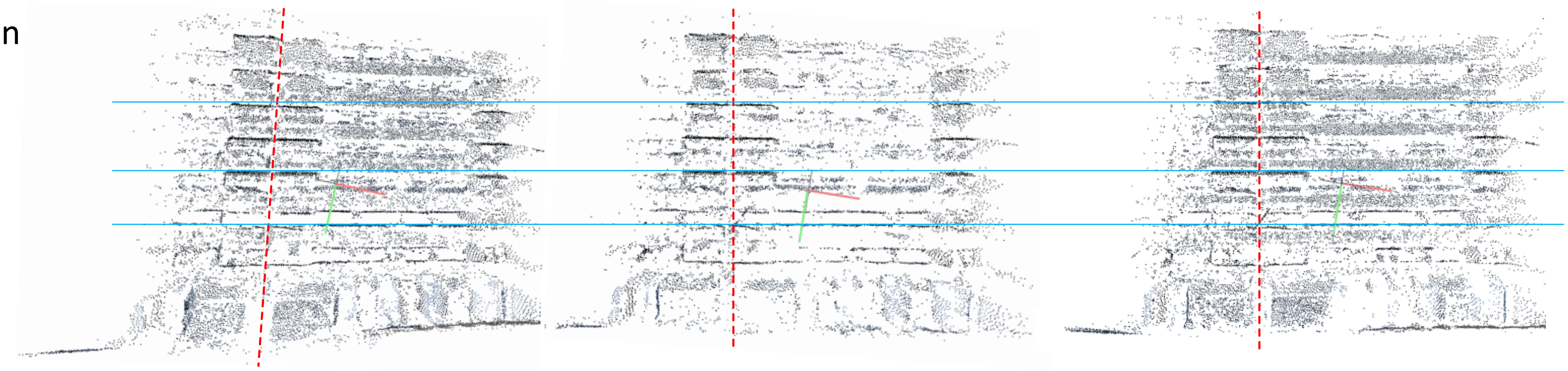


[1] Forssen P E , Ringaby E. Rectifying rolling shutter video from hand-held devices. CVPR 2010.

2.2. SUNet: Symmetric Undistortion Network for RS Correction

Experiments

- 3D reconstruction



(a) Original (b) Corrected (c) Ground Truth

- Inference time

Method	Time	Hardware
DiffSfM (SOTA classic-model-based)	~ 8 minutes	i7-7700K CPU
DeepUnrollNet (SOTA deep-learning-based)	0.34 seconds	NVIDIA GeForce 2080Ti GPU
SUNet (Ours)	**0.21 seconds**	NVIDIA GeForce 2080Ti GPU
	640×480 image resolution	

2.2. SUNet: Symmetric Undistortion Network for RS Correction

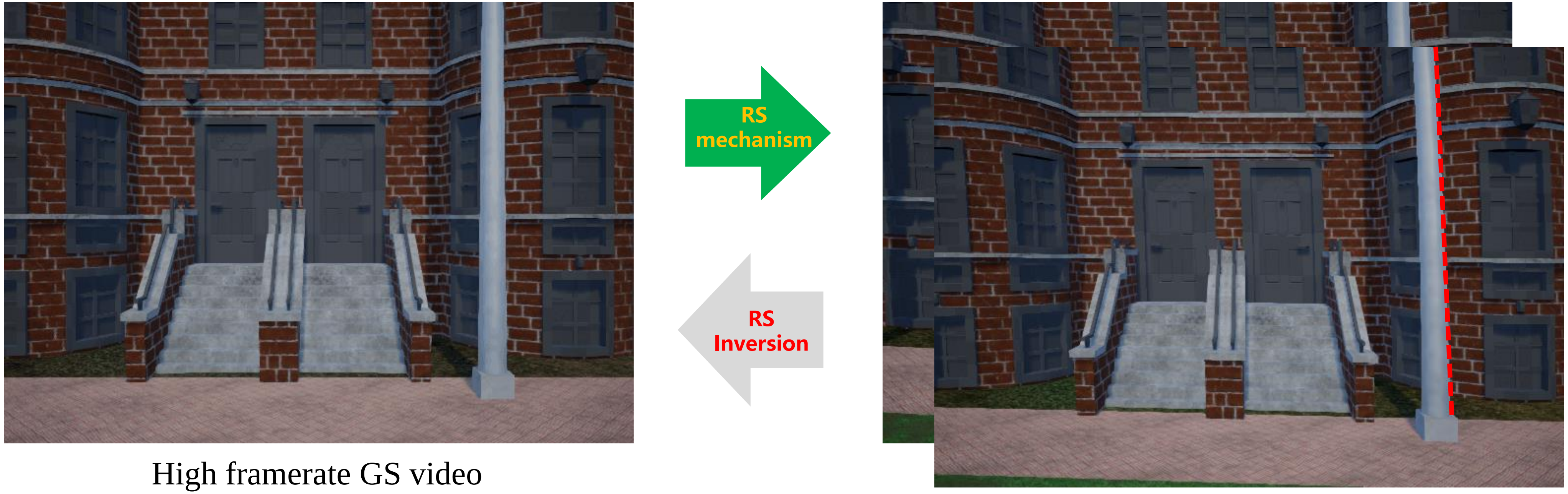
Conclusions

- Recovering GS image corresponding to the exposure time of the first scanline is of both theoretical interest and great practical importance, such as multi-sensor fusion, computational photography, autonomous driving, etc.
- The observation that **the first and second RS images have different contributions to different regions of the target GS image** is fundamental and helpful. And the idea of promoting this property through a symmetry consistency constraint is reasonable.
- A distinct advantage of the method is the use of **symmetric network architecture** to improve the efficient aggregation of contextual information.
- The **context-aware cost volume** we construct can effectively promote contextual consistency at different scales.
- Extensive experiments demonstrate that our approach performs favorably against the state-of-the-art methods in both GS image restoration and inference efficiency.
- Maybe applicable to other frame interpolation tasks.

3. Rolling Shutter Temporal Super-Resolution

Background:

➤ Rolling shutter images can be viewed as the result of the **row-wise combination** of global shutter images captured by a virtual moving GS camera over the period of camera readout time.



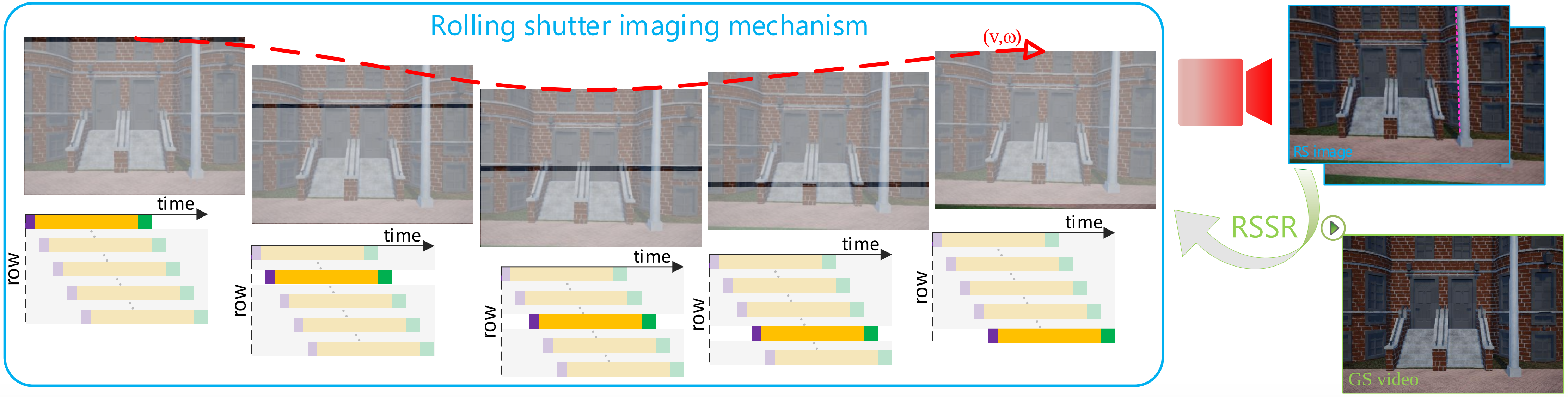
High framerate GS video

Two consecutive RS frames

3. Rolling Shutter Temporal Super-Resolution

Objective:

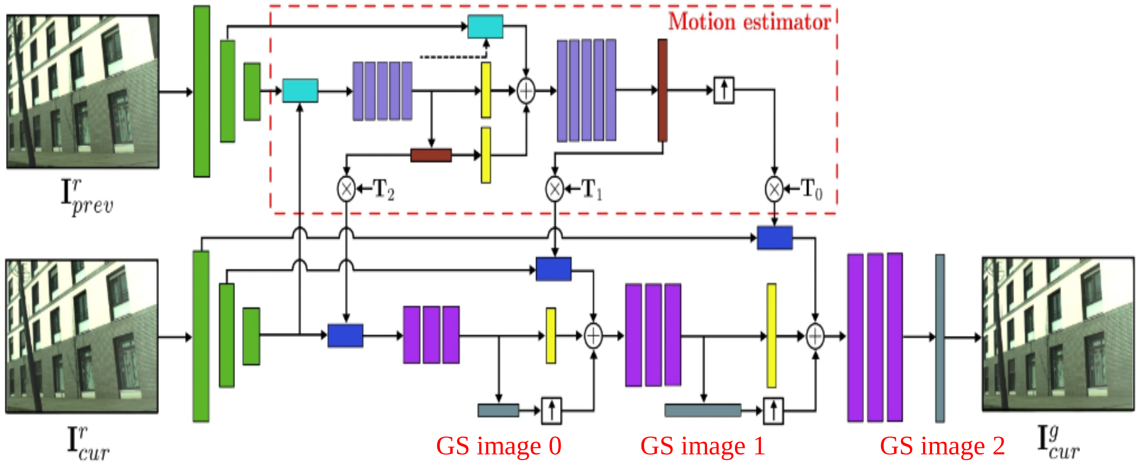
- Invert the rolling shutter imaging mechanism, i.e., **RS temporal super-resolution (RSSR),** is extremely challenging, e.g., recovering 960 GS images from two 480-height RS images, **which is far from being solved in the deep learning framework**.
- Different from estimating the undistortion flow at a specific time in RS correction, here it is necessary to **estimate the undistortion flow at any time**. Therefore, it is crucial to build connections between them.



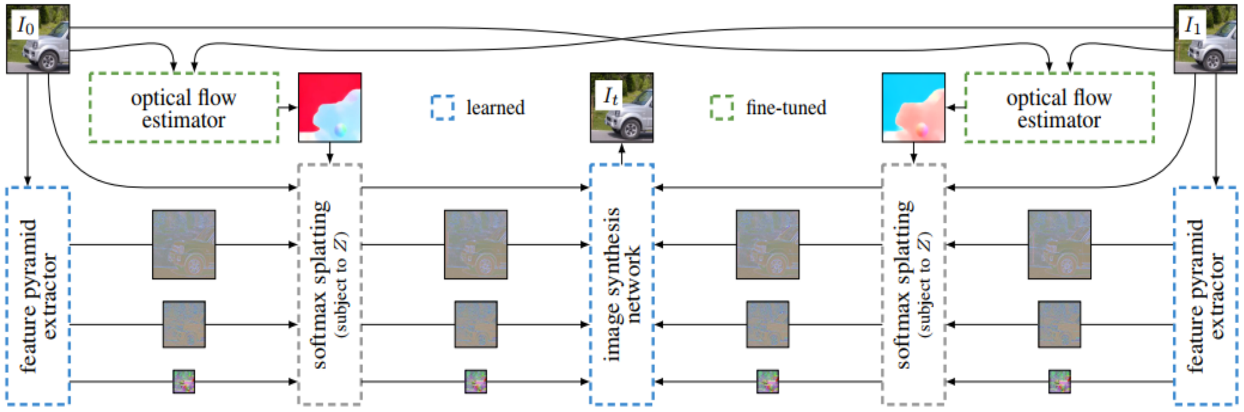
3. Rolling Shutter Temporal Super-Resolution

Challenges:

- Beyond eliminating the geometric RS distortion in **the two-view RS correction task**, we have to output a high framerate GS image sequence as well as ensure its temporal smoothness.
- Different from the slight and controllable pixel displacement in **the GS video interpolation task**, which is located inside its optical flow, the pixel displacement when correcting the RS image may exceed its local neighborhood defined by its optical flow.



SOTA two-view RS correction method [Liu et al, CVPR'20]: only one reliable GS image can be recovered.



SOTA GS video interpolation method [Niklaus et al, CVPR'20]: is incapable of reducing the RS artifacts.

Inverting a Rolling Shutter Camera: Bring Rolling Shutter Images to High Framerate Global Shutter Video

*Bin Fan, Yuchao Dai**

ICCV 2021

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Contributions:

- We identify and establish a detailed proof of the scanline-dependent nature of the bidirectional undistortion flows, which is essential for understanding the intrinsic geometrical properties of RS correction problem.
- From the theoretical perspective, we propose **the first geometry-aware learning-based RSSR solution** for latent GS video sequence extraction from two consecutive RS images, which brings RS images alive.
- Our approach not only outperforms the state-of-the-art methods in both RS effect removal and inference efficiency, but also can produce a smooth and continuous GS video.

Input RS image 1 Input RS image 2 Output GS video

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Differential RS Geometry

- GS-aware forward warping:

$$\mathbf{f} = \frac{\mathbf{A}\mathbf{v}}{Z} + \mathbf{B}\boldsymbol{\omega} \triangleq \pi(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f), \qquad (1)$$

where

$$\mathbf{A} = \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix},$$
$$\mathbf{B} = \begin{bmatrix} \frac{xy}{f} & -\left(f + \frac{x^2}{f}\right) & y \\ \left(f + \frac{y^2}{f}\right) & -\frac{xy}{f} & -x \end{bmatrix}. \qquad (2)$$

Here, (x, y) is the normalized image coordinate and f denotes the focal length.

- RS-aware warping:

(**Optical flow between two consecutive RS images**)

$$\begin{bmatrix} \mathbf{f}_u \\ \mathbf{f}_v \end{bmatrix} = \alpha \begin{bmatrix} \pi_u(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \\ \pi_v(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \end{bmatrix},$$

where

$$\alpha = 1 + \frac{\gamma \boxed{\mathbf{f}_v}}{h}$$

constant velocity motion model[1]

α is the RS-aware interpolation factor, depending on the corresponding RS optical flow; γ is the readout time ratio.

- Note that we prove that γ is positive for forward RS optical flow (i.e. from frame 1 to frame 2) and γ is negative for backward RS optical flow (i.e. from frame 2 to frame 1) .
- Try to eliminate $\boxed{\mathbf{f}_v}$ …… (later)

[1] Zhuang B, Cheong L F, Hee Lee G. Rolling-shutter-aware differential sfm and image rectification. ICCV 2017.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Undistortion Flow vs Optical Flow

➢ **Bidirectional RS undistortion flow**:

$$\begin{bmatrix} \mathbf{u}_u \\ \mathbf{u}_v \end{bmatrix} = \beta \begin{bmatrix} \pi_u(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \\ \pi_v(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \end{bmatrix},$$

where

$$\beta = \frac{\gamma(s-\kappa)}{h}$$

which delivers each RS pixel $\mathbf{x}$ on κ-th scanline to its GS canvas defined by the pose corresponding to s-th scanline.

➢ **Mutual conversion between varying RS undistortion flows that correspond to different scanlines**:

$$\begin{bmatrix} \mathbf{u}_u^{s_2} \\ \mathbf{u}_v^{s_2} \end{bmatrix} = \frac{s_2-\kappa}{s_1-\kappa} \begin{bmatrix} \mathbf{u}_u^{s_1} \\ \mathbf{u}_v^{s_1} \end{bmatrix}$$

Assuming that two GS images corresponding to s_1-th scanline and s_2-th scanline are to be restored.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Undistortion Flow vs Optical Flow

- **Bidirectional RS undistortion flow**:

$$\begin{bmatrix} \mathbf{u}_u \\ \mathbf{u}_v \end{bmatrix} = \beta \begin{bmatrix} \pi_u(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \\ \pi_v(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \end{bmatrix},$$

where

$$\beta = \frac{\gamma(s-\kappa)}{h}$$

which delivers each RS pixel $\mathbf{x}$ on κ-th scanline to its GS canvas defined by the pose corresponding to s-th scanline.

- **Bidirectional RS optical flow**: (By eliminating $\mathbf{f}_v$)

$$\begin{bmatrix} \mathbf{f}_u \\ \mathbf{f}_v \end{bmatrix} = \frac{h}{h-\gamma\pi_v} \begin{bmatrix} \pi_u \\ \pi_v \end{bmatrix}$$

which models the transformation of RS pixel $\mathbf{x}$ between two consecutive RS frames.

- **Connection between the undistortion flow and optical flow**:

$$\begin{bmatrix} \mathbf{u}_u \\ \mathbf{u}_v \end{bmatrix} = c \begin{bmatrix} \mathbf{f}_u \\ \mathbf{f}_v \end{bmatrix},$$

where

$$c = \frac{\gamma(s-\kappa)(h-\gamma\pi_v)}{h^2}$$

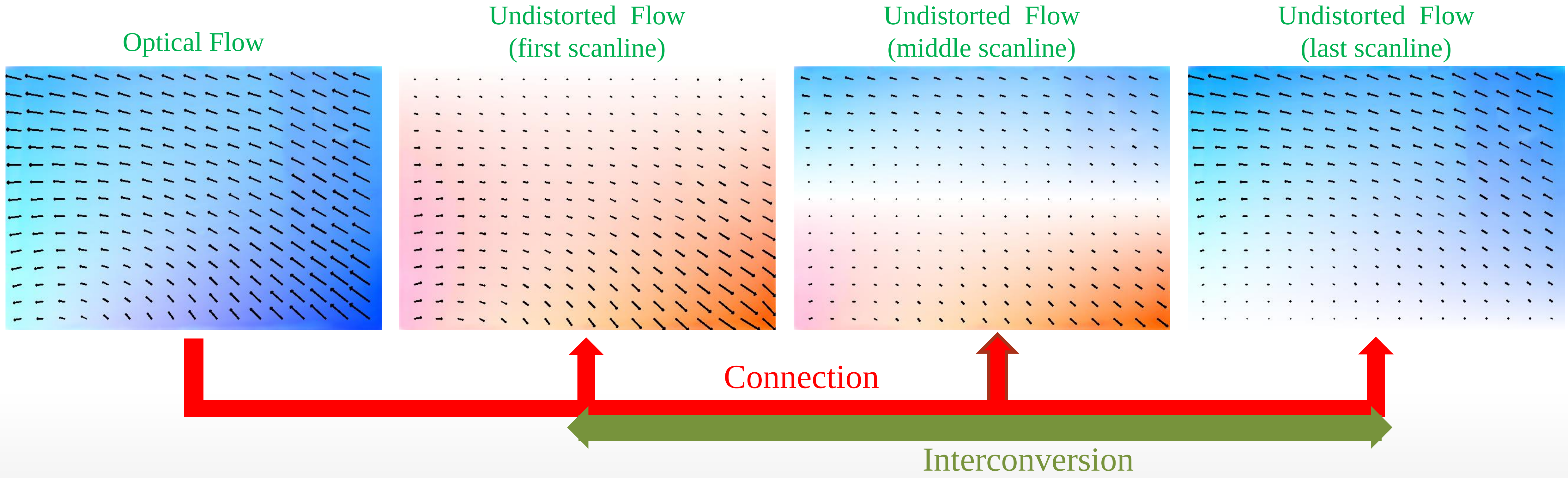
Correlation factor

- Note that we prove $c \in (-1,1)$ when correcting an RS image to its middle-scanline GS image.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Undistortion Flow vs Optical Flow

- The undistortion flows exhibit a more significant **scanline dependence**.
- The undistortion flows near the target scanline appear as smaller warping displacement values;
  The undistortion flows corresponding to pixels that are opposite to the target scanline show different warping displacement directions.



3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Constant Velocity Propagation vs. Constant Acceleration Propagation:

- **Undistortion flow (Constant Velocity Model):**

$$\begin{bmatrix} \mathbf{u}_u \\ \mathbf{u}_v \end{bmatrix} = \beta \begin{bmatrix} \pi_u(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \\ \pi_v(\mathbf{v}, \boldsymbol{\omega}, \mathbf{x}, Z, f) \end{bmatrix},$$

where

$$\beta = \frac{\gamma(s-\kappa)}{h}$$

which delivers each RS pixel $\mathbf{x}$ on κ-th scanline to its GS canvas defined by the pose corresponding to s-th scanline.

- **Mutual conversion between varying undistortion flows that correspond to different scanlines (Constant Velocity Model):**

$$\begin{bmatrix} \mathbf{u}_u^{s_2} \\ \mathbf{u}_v^{s_2} \end{bmatrix} = \frac{s_2-\kappa}{s_1-\kappa} \begin{bmatrix} \mathbf{u}_u^{s_1} \\ \mathbf{u}_v^{s_1} \end{bmatrix}$$

Assuming that two GS images corresponding to s_1-th scanline and s_2-th scanline are to be restored.

- **Undistortion flow (Constant Acceleration Model):**

$$\beta = \frac{\gamma(s-\kappa)}{h} \cdot \frac{2h + k\gamma(s-\kappa)}{h(k+2)}$$

- **Mutual conversion between varying undistortion flows that correspond to different scanlines (Constant Acceleration Model):**

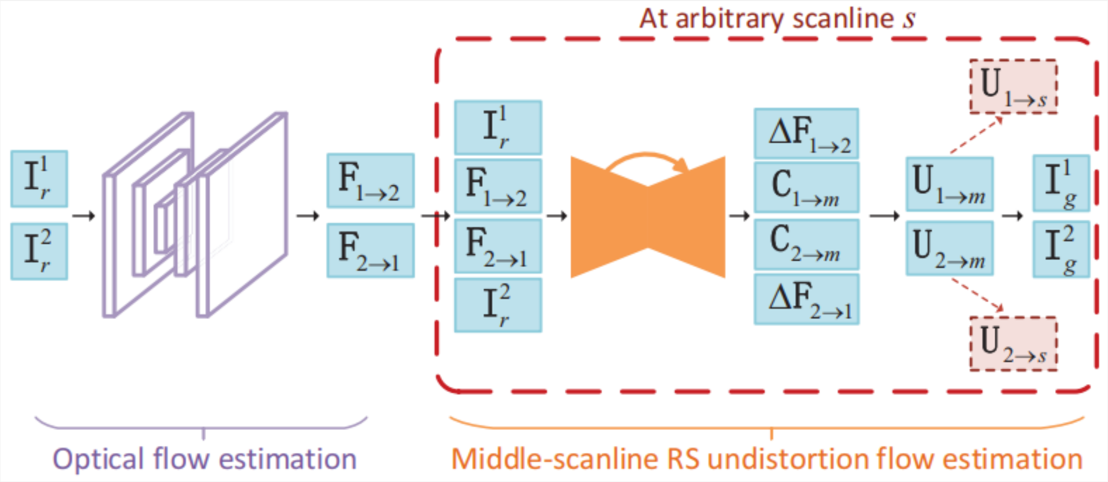
$$\begin{bmatrix} \mathbf{u}_u^{s_2} \\ \mathbf{u}_v^{s_2} \end{bmatrix} = \frac{(s_2-\kappa)(2h + \boxed{k\gamma}(s_2-\kappa))}{(s_1-\kappa)(2h + \boxed{k\gamma}(s_1-\kappa))} \begin{bmatrix} \mathbf{u}_u^{s_1} \\ \mathbf{u}_v^{s_1} \end{bmatrix}$$

Fan B, Dai Y, Li H. Rolling shutter inversion: bring rolling shutter images to high framerate global shutter video. IEEE TPAMI 2022.

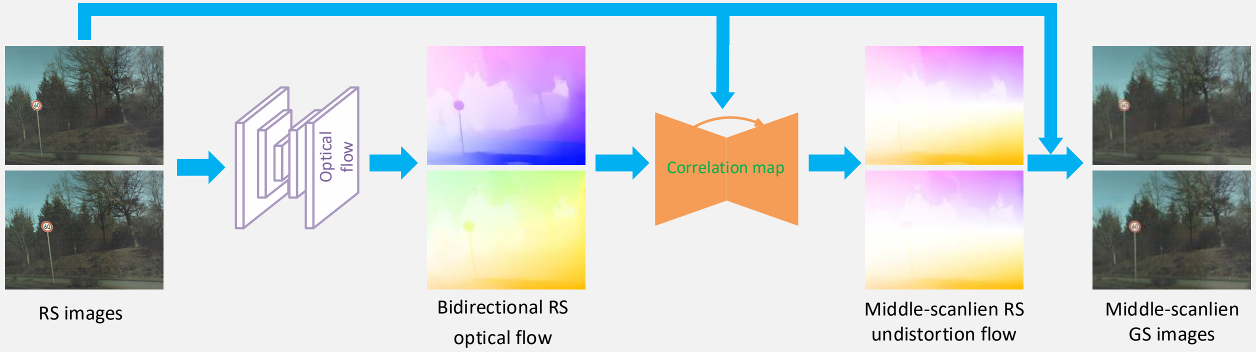
3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Pipeline of our RSSR method:

- Firstly, we estimate the **bidirectional optical flows** by using the classic PWC-Net.
- Secondly, we use a UNet network to learn the **middle-scanline correlation map** such that the middle-scanline undistortion flows can be inferred. Meanwhile, **undistortion flows** for any scanline can be associated and propagated explicitly.
- Finally, the **forward warping** is employed to warp RS images, yielding a GS video sequence corresponding to arbitrary scanlines.



$$\mathbf{U}_{1\to m} = \mathbf{C}_{1\to m} \odot (\mathbf{F}_{1\to2} + \Delta\mathbf{F}_{1\to2}),$$
$$\mathbf{U}_{2\to m} = \mathbf{C}_{2\to m} \odot (\mathbf{F}_{2\to1} + \Delta\mathbf{F}_{2\to1}).$$



- ***Only the middle-scanline GS images are needed for supervision.***

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Quantitative Results

Table 1: Quantitative comparisons on recovering GS images corresponding to the first scanline of the second RS frame. The numbers in red and blue represent the best and second-best performance. Note that we cannot benchmark the Fastec-RS dataset due to its lack of training ground truth. Regardless of the black edges of corrected images, our approach performs favorably against other methods.

First scanline:

Method	PSNR↑			SSIM↑		LPIPS↓	
	CRM	CR	FR	CR	FR	CR	FR
DeepUnrollNet [18]	26.90	26.46	26.52	0.81	0.79	0.0703	0.1222
DiffHomo [37]	19.60	18.94	18.68	0.61	0.61	0.1798	0.2229
DiffSfM-*PWCNet* [36]	19.53	18.62	18.59	0.69	0.63	0.2042	0.2416
DiffSfM-*RAFT* [36]	24.20	21.28	20.14	0.78	0.70	0.1322	0.1789
RSSR (Ours)	30.17	24.78	21.26	0.87	0.78	0.0695	0.1424

Table 3: Quantitative comparisons of the performance between our approach and DeepUnrollNet [4] in recovering GS images corresponding to the middle scanline of the second RS frame. Note that, in other chapters and the main manuscript, all competing methods refer to the first scanline of the second RS frame.

Middle scanline:

Method	PSNR↑			SSIM↑		LPIPS↓	
	CRM	CR	FR	CR	FR	CR	FR
DeepUnrollNet [4]	27.86	**27.54**	**27.02**	0.829	0.828	0.0555	**0.0791**
RSSR (Ours)	**29.36**	26.57	25.01	**0.900**	**0.834**	**0.0553**	0.0817

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

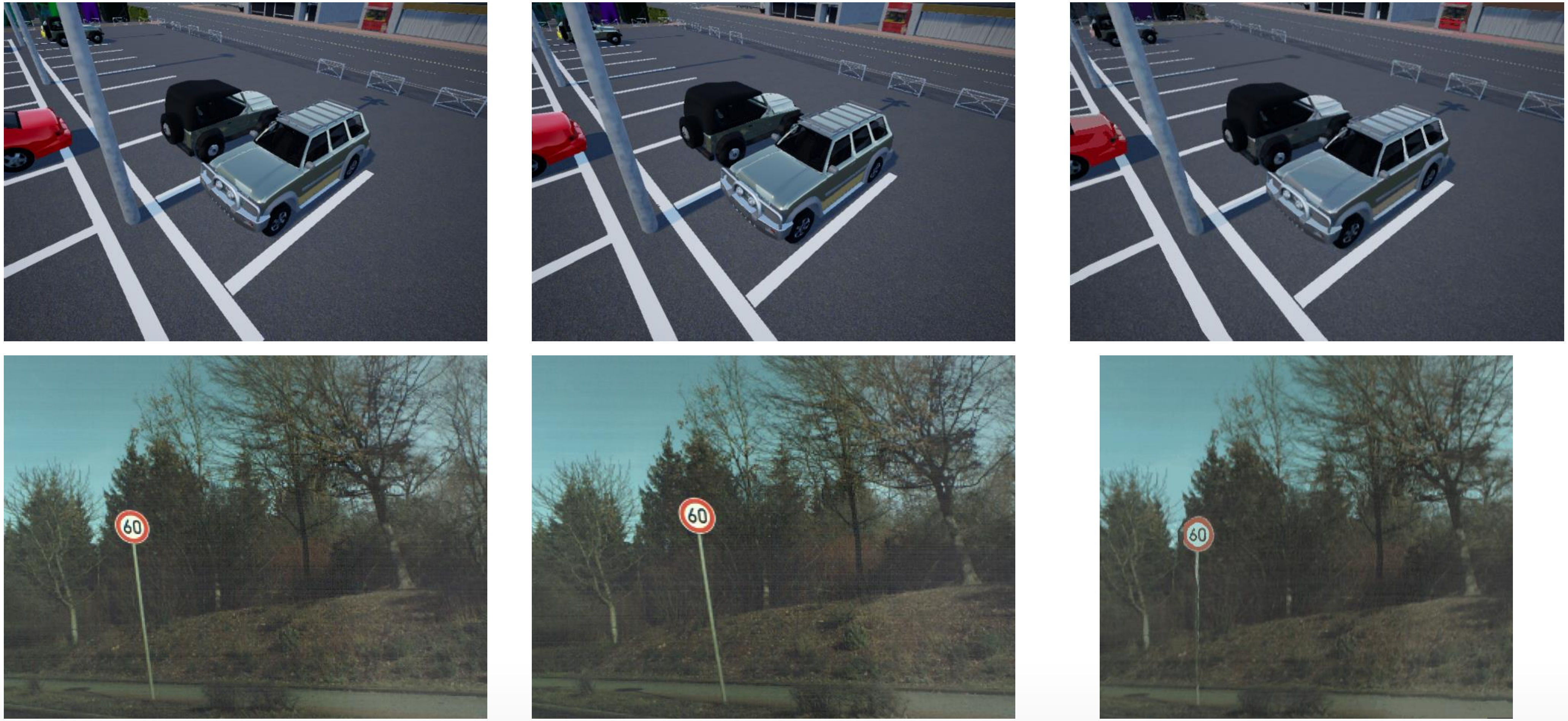
- Qualitative Results

Inputs (Overlay) | DeepUnrollNet | DiffHomo | DiffSfM | RSSR (Ours) | GT

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Generating high framerate GS videos (synthetic data[1])



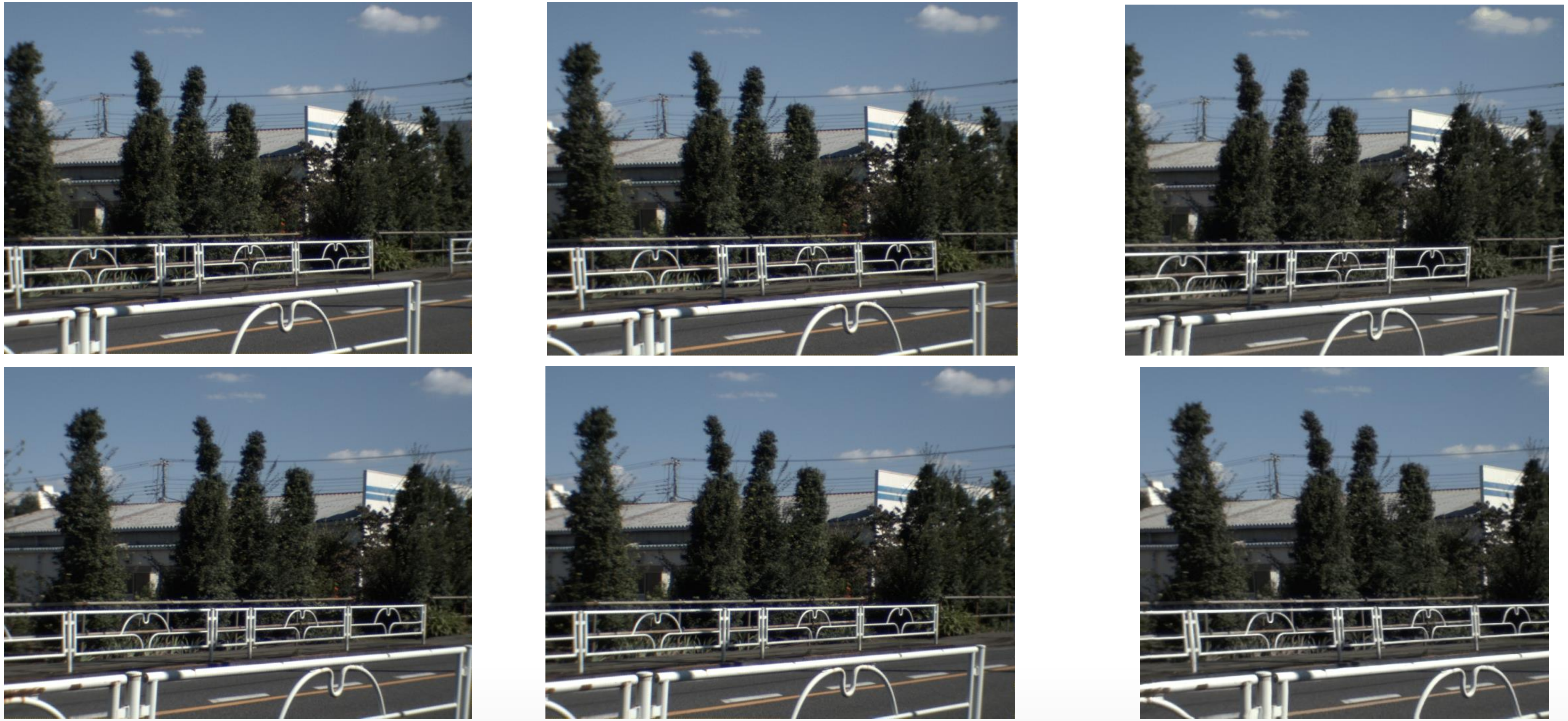
Input RS Frame 1 | Input RS Frame 2 | Our RSSR Result (Cropped)

[1] Liu P, Cui Z, Larsson V, et al. Deep shutter unrolling network. CVPR 2020.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Generating high framerate GS videos (Real data[1] by an RS camera mounted on a car)



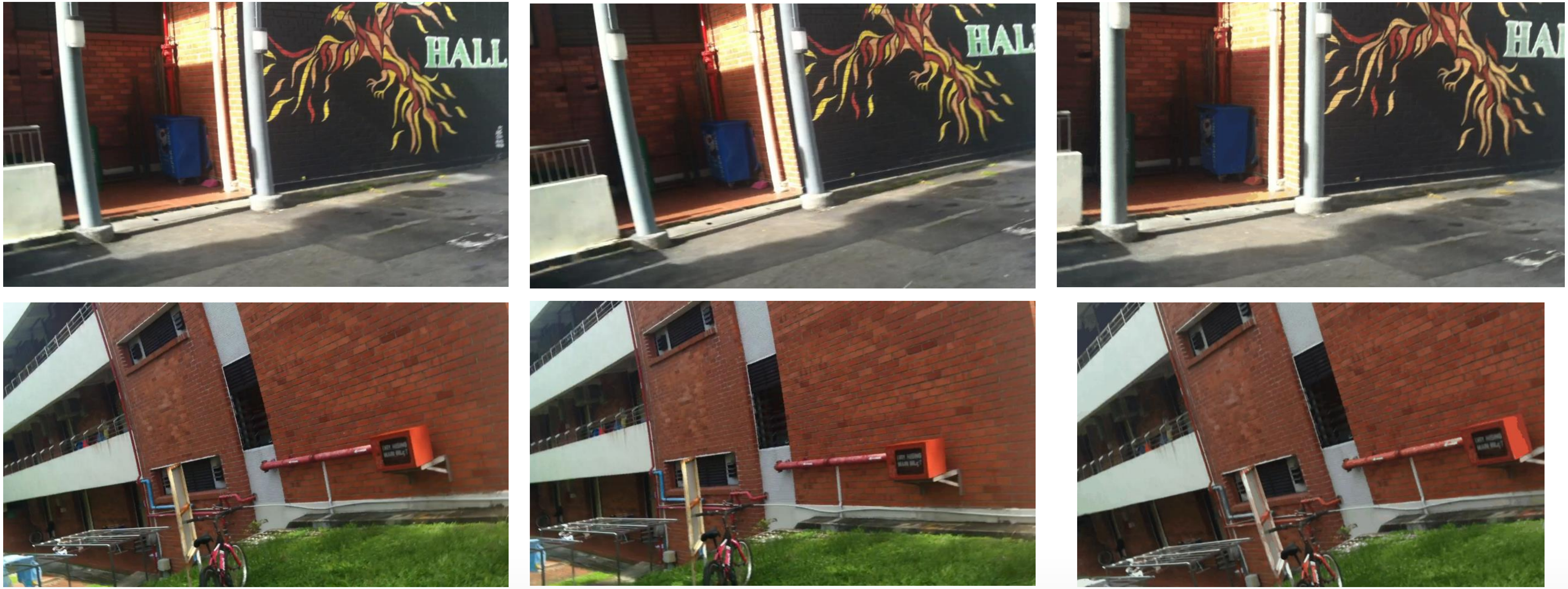
Input RS Frame 1 | Input RS Frame 2 | Our RSSR Result (Cropped)

[1] Cao M, Zhong Z, Wang J, et al. Learning adaptive warping for real-world rolling shutter correction. CVPR 2022.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Generating high framerate GS videos (real data[1])



Input RS Frame 1 | Input RS Frame 2 | Our RSSR Result (Cropped)

[1] Zhuang B, Cheong L F, Hee Lee G. Rolling-shutter-aware differential sfm and image rectification. ICCV 2017.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Comparison with SOTA video frame interpolation methods

Inputs (Overlay) | BMBC | DAIN | RSSR (Ours) | GT

Figure 3: Visual results against video frame interpolation algorithms (BMBC [6] and DAIN [1]) to generate an intermediate frame corresponding to the intermediate time of two consecutive RS frames. Only our proposed RSSR method can successfully remove RS artifacts.

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Experiments

- Comparison with two-stage method (baseline)

Given *three* consecutive RS images, we first obtain two corrected GS images in sequence by using DeepUnrollNet [CVPR'20], and then interpolate the GS image corresponding to the first scanline of the third RS image using DAIN [CVPR'19].

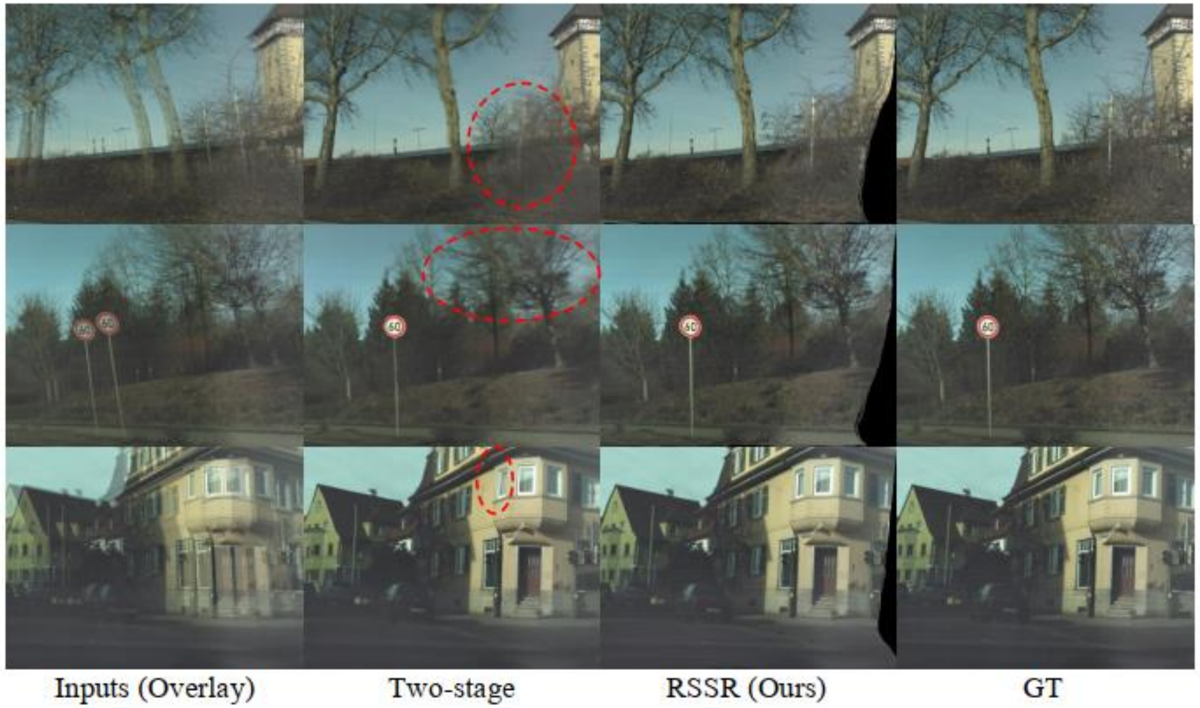


Figure 4: Visual results against the two-stage approach: perform RS correction first, then perform video frame interpolation.

- Inference times

Methods	Times	Outputs
DeepUnrollNet (SOTA)	0.34 s	1 GS image
Two-stage method	5 min	960 GS images
RSSR (Ours)	**0.12 s**	2 GS images
	1.8 s	960 GS images

Test on an NVIDIA GeForce RTX 2080Ti GPU with 640×480 image resolution

3.1. Inverting a RS Camera: Bring RS Images to High Framerate GS Video

Conclusions

- We have revealed the intrinsic geometrical properties of RS correction problem and made **three contributions**: 1) formulating the bidirectional RS undistortion flows under the constant velocity motion model, 2) building the connection between the RS undistortion flow and optical flow via a scaling operation, and 3) developing a mutual conversion scheme between varying RS undistortion flows that correspond to different scanlines.
- We have proposed **the first geometry-aware learning-based RSSR solution** for latent GS video sequence extraction from two consecutive RS images, which brings RS images alive.
- Our rolling shutter temporal super-resolution pipeline marries the advantage of RS geometric reasoning and modern deep learning-enabled computer vision, which can effectively explore the underlying spatio-temporal geometric relationships.
- Extensive experiments demonstrate that our approach achieves joint RS correction and temporal super-resolution, outperforming state-of-the-art methods.
- Our preliminary implementation can very efficiently **generate 960 GS images with 640×480 resolution in 1.8 seconds** on an NVIDIA 2080Ti GPU.

Context-Aware Video Reconstruction for Rolling Shutter Cameras

Bin Fan, Yuchao Dai, Zhiyuan Zhang, Qi Liu, Mingyi He*

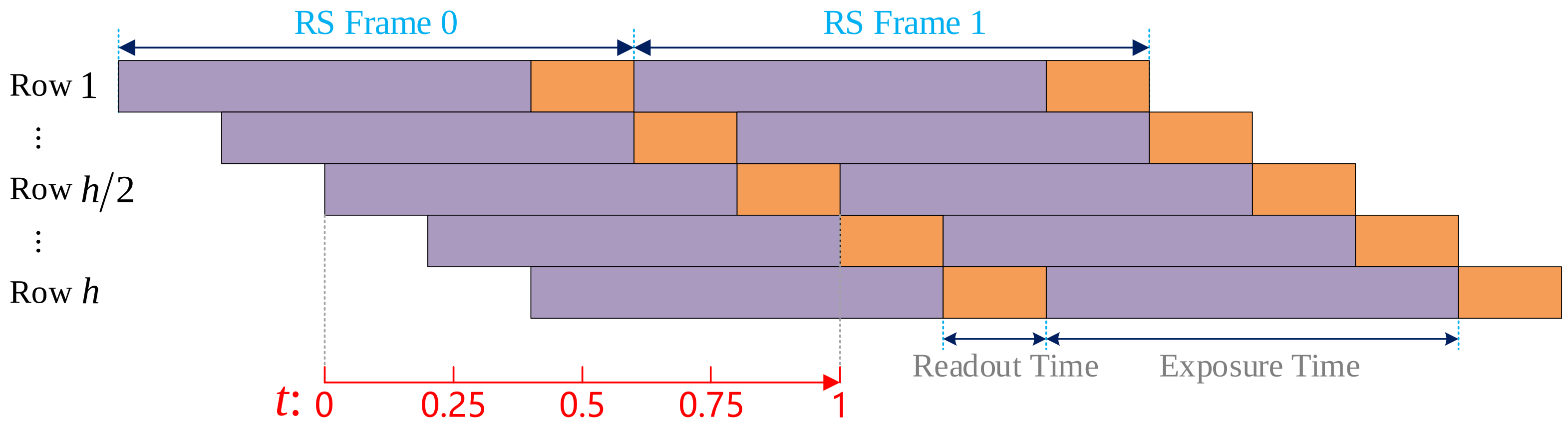
CVPR 2022

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Formulation:

Inspired by the task of video frame interpolation, we re-define the RS temporal super-resolution problem **in the temporal dimension**.

- Given two RS frames at adjacent times 0 and 1, we aim to synthesize an intermediate GS frame corresponding to any time t, where $0 \leq t \leq 1$.
- In particular, the middle scanlines of the two RS images correspond to time instances 0 and 1, respectively.



3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

□ Motivation:

The geometry-aware RS inversion proposed in ICCV 2021 warps RS frames directly.

RS frame 0 t=0 t=0.25 t=0.5 t=0.75 t=1

Hence, it suffers from two limitations:

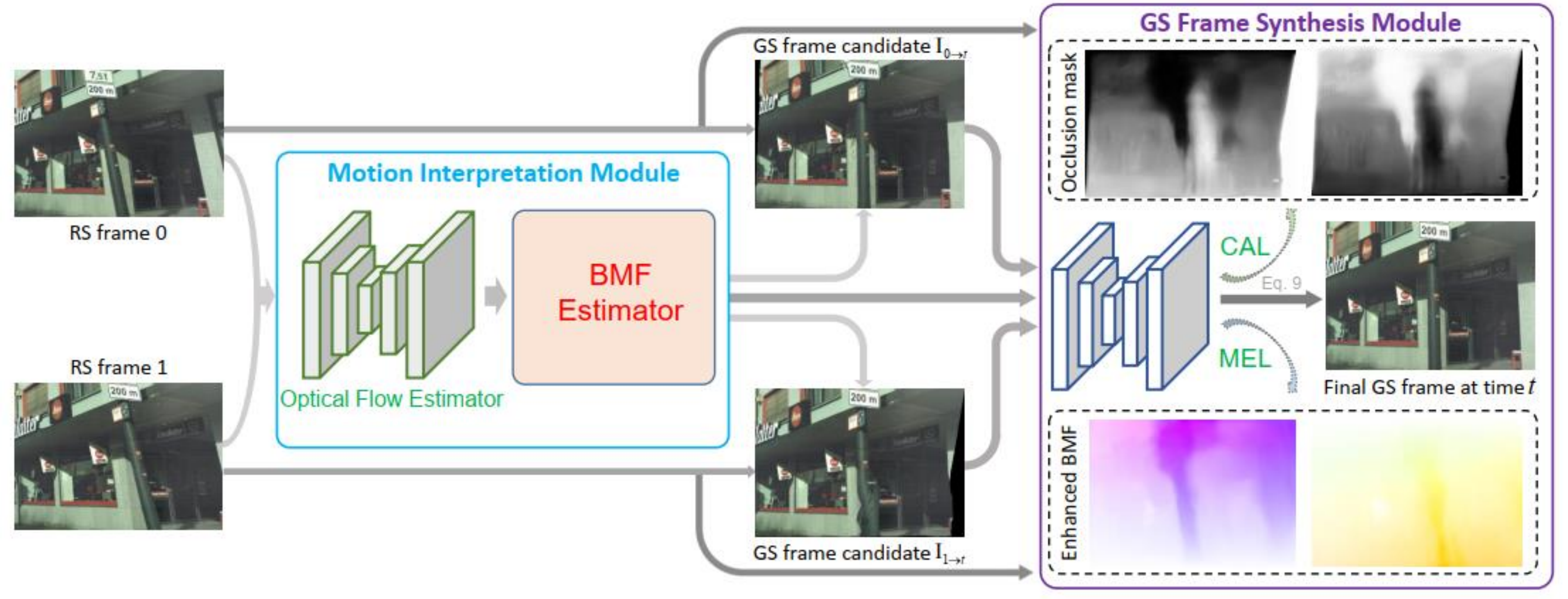
✓ **Masses of black holes.** This is a common issue for warping-based methods due to the occlusion. To maintain visual consistency, a cropping operation is used to discard the holes, but may degrade the visual experience.

✓ **Noticeable object-specific motion artifacts**. When recording dynamic scenes, the moving objects violate the assumption of constant velocity motion, resulting in severe motion artifacts.

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Pipeline of our CVR method:

1. **Motion interpretation module**, including a network-based bilateral motion field estimator (NBMF) or an approximated bilateral motion field estimator (ABMF).
2. **GS frame synthesis module**, including a motion enhancement layer (MEL) and a contextual aggregation layer (CAL).



3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Details of motion interpretation module:

- The bilateral motion field can be generated by scaling the regular optical flow field, i.e.,

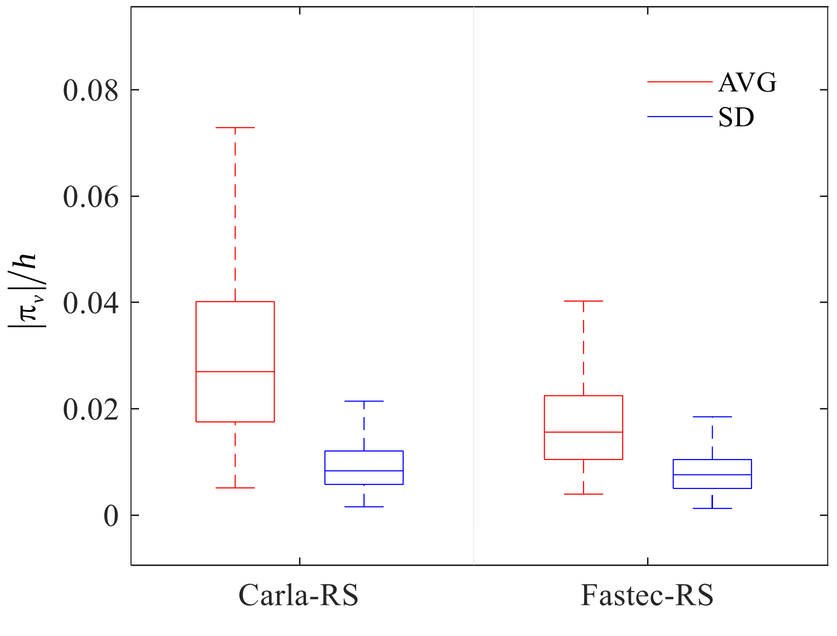
$$\mathbf{U}_{0\to t}(\mathbf{x}) = \mathbf{C}_{0\to t}(\mathbf{x}) \cdot \mathbf{F}_{0\to 1}(\mathbf{x})$$
$$\mathbf{U}_{1\to t}(\mathbf{x}) = \mathbf{C}_{1\to t}(\mathbf{x}) \cdot \mathbf{F}_{1\to 0}(\mathbf{x})$$

- The bilateral correction map was formulated under the constant camera motion in ICCV 2021, which can be learned by an encoder-decoder network.

$$\mathbf{C}_{0\to t}(\mathbf{x}) = \frac{(t-\tau_0)(h-\pi_v)}{h}$$
$$\mathbf{C}_{1\to t}(\mathbf{x}) = \frac{(\tau_1-t)(h+\pi_v')}{h}$$

- In this work, we further propose its approximated version neglecting the parallax effects, which is independent of image content and can be pre-defined.

$$\mathbf{C}_{0\to t}(\mathbf{x}) = t-\tau_0$$
$$\mathbf{C}_{1\to t}(\mathbf{x}) = \tau_1-t$$



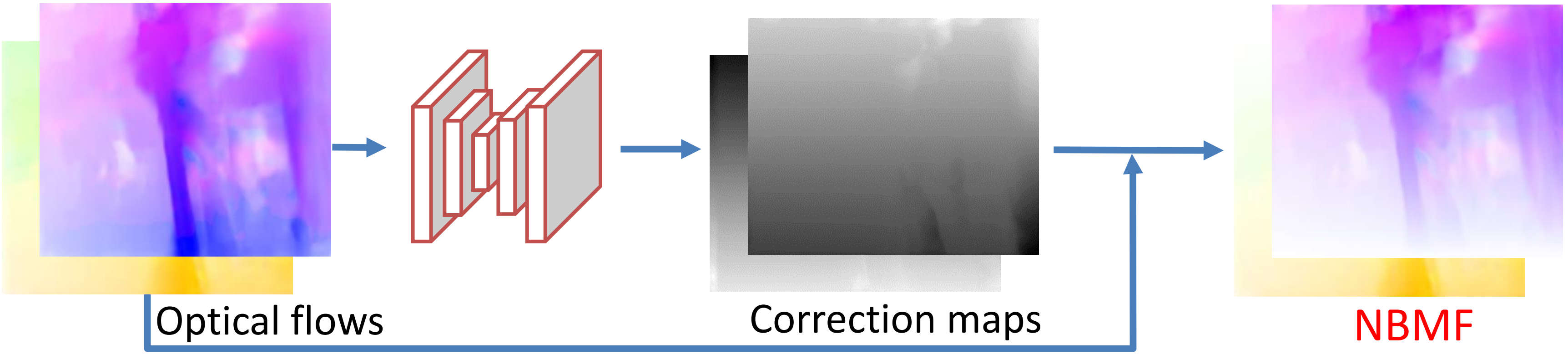
3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Details of motion interpretation module:

1. Bidirectional optical flow estimator.
2. Bilateral motion field estimator (NBMF or ABMF) at arbitrary time t∈[0,1].

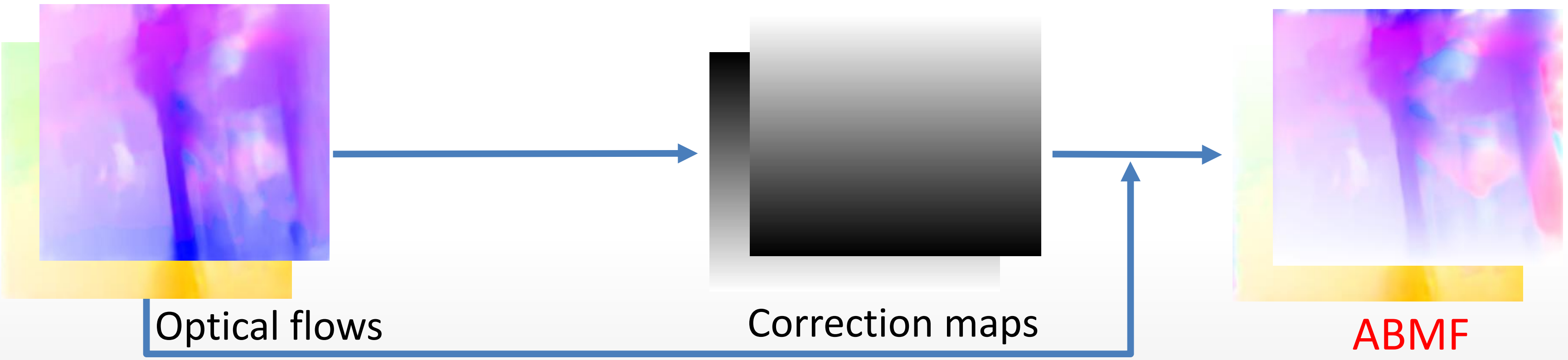
- **Network-based BMF (NBMF)**

$$\mathbf{C}_{0\to t}(\mathbf{x}) = \frac{(t-\tau_0)(h-\pi_v)}{h}$$
$$\mathbf{C}_{1\to t}(\mathbf{x}) = \frac{(\tau_1-t)(h+\pi_v')}{h}$$



- **Approximated BMF (ABMF)**

$$\mathbf{C}_{0\to t}(\mathbf{x}) = t-\tau_0$$
$$\mathbf{C}_{1\to t}(\mathbf{x}) = \tau_1-t$$

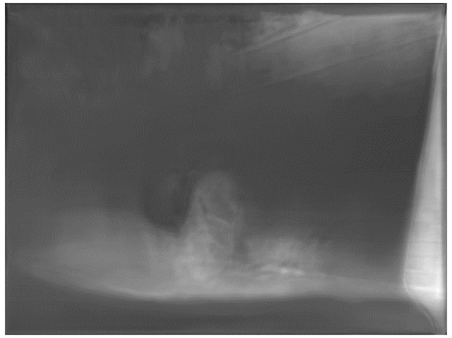


3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

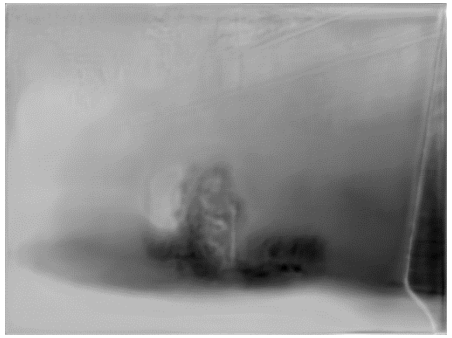
Details of GS frame synthesis module:

- BMF residuals are estimated to improve the final flow quality in boundaries and unsmooth regions.

$$\hat{\mathbf{U}}_{0\to t} = \mathbf{U}_{0\to t} + \Delta\mathbf{U}_{0\to t}$$
$$\hat{\mathbf{U}}_{1\to t} = \mathbf{U}_{1\to t} + \Delta\mathbf{U}_{1\to t}$$



$\|\Delta\mathbf{U}_{0\to 0.5}\|_2$



$\|\Delta\mathbf{U}_{1\to 0.5}\|_2$

- Bilateral occlusion masks are generated to guide GS frame synthesis to handle occlusions.

$$\hat{\mathbf{I}}_t^g = \frac{(1-t)\mathbf{O}_{0\to t}\hat{\mathbf{I}}_{0\to t}^g + t\mathbf{O}_{1\to t}\hat{\mathbf{I}}_{1\to t}^g}{(1-t)\mathbf{O}_{0\to t} + t\mathbf{O}_{1\to t}}$$

$\mathbf{O}_{0\to 0.5}$

$\mathbf{O}_{1\to 0.5}$

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Loss functions:

- Reconstruction loss: $\mathcal{L}_r = \frac{1}{T}\sum_{i=1}^{T}\left\|\hat{\mathbf{I}}_{t_i}^{g} - \mathbf{I}_{t_i}^{gt}\right\|_1$

- Perceptual loss: $\mathcal{L}_p = \frac{1}{T}\sum_{i=1}^{T}\left\|\phi\left(\hat{\mathbf{I}}_{t_i}^{g}\right) - \phi\left(\mathbf{I}_{t_i}^{gt}\right)\right\|_1$

- Contextual consistency loss: $\mathcal{L}_c = \frac{1}{2T}\sum_{i=1}^{T}\left(\left\|\hat{\mathbf{I}}_{0\to t_i}^{g} - \mathbf{I}_{t_i}^{gt}\right\|_1 + \left\|\hat{\mathbf{I}}_{1\to t_i}^{g} - \mathbf{I}_{t_i}^{gt}\right\|_1\right)$

- Total variation loss: $\mathcal{L}_{tv} = \frac{1}{2T}\sum_{i=1}^{T}\left(\left\|\nabla\hat{\mathbf{U}}_{0\to t_i}\right\|_2 + \left\|\nabla\hat{\mathbf{U}}_{1\to t_i}\right\|_2\right)$

Note that we use the ground-truth GS images corresponding to times 0.0, 0.5 and 1.0 to supervise the network training.

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Quantitative results of RS effect removal

Time t=0.5:

Table 1. Quantitative comparisons on recovering GS images at time step $t = 0.5$. The numbers in **red** and blue represent the best and second-best performance. Our method is far superior to baseline methods and the proposed ABMF model is effective as an initialization.

Method	Runtime (seconds)	PSNR↑ (dB)			SSIM↑		LPIPS↓	
		CRM	CR	FR	CR	FR	CR	FR
DiffSfM [62]	467	24.20	21.28	20.14	0.775	0.701	0.1322	0.1789
DiffHomo [63]	424	19.60	18.94	18.68	0.606	0.609	0.1798	0.2229
DeepUnrollNet [24]	0.34	26.90	26.46	26.52	0.807	0.792	0.0703	0.1222
SUNet [10]	0.21	29.28	29.18	28.34	0.850	0.837	0.0658	0.1205
RSSR*	0.09	28.20	23.86	21.02	0.839	0.768	0.0764	0.1866
RSSR [9]	0.12	30.17	24.78	21.23	0.867	0.776	0.0695	0.1659
*CVR** (Ours)	0.12	31.82	31.60	28.62	0.927	0.845	0.0372	0.1117
CVR (Ours)	0.14	**32.02**	**31.74**	**28.72**	**0.929**	**0.847**	**0.0368**	**0.1107**

*: *applying our proposed approximated bilateral motion field (ABMF) model.*

Time t=1.0:

Table A1. Quantitative comparisons on recovering GS images at time step $t = 1$. The numbers in **red** and blue represent the best and second-best performance. In addition to the SOTA quantification performance for GS image recovery at time $t = 0.5$, our method also obtains almost consistent best metrics at time $t = 1$. Note that not only these, high-quality GS video frames corresponding to any time $t \in [0, 1]$ can be accurately estimated by our method.

Method	PSNR↑ (dB)			SSIM↑		LPIPS↓	
	CRM	CR	FR	CR	FR	CR	FR
DeepUnrollNet [24]	27.86	27.54	**27.02**	0.829	0.828	0.0555	**0.0791**
RSCD [61]	-	-	24.84	-	0.778	-	0.1070
RSSR [9]	29.36	26.57	24.89	0.900	0.824	0.0553	0.1109
*CVR** (Ours)	28.28	28.19	26.58	0.912	0.833	0.0444	0.1014
CVR (Ours)	**29.41**	**29.19**	26.67	**0.915**	**0.838**	**0.0403**	0.1011

*: *applying our proposed approximated bilateral motion field (ABMF) model.*

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Quantitative results of RS effect removal

Input RS Frame 0

SUNet (ICCV 2021)

DiffHomo (ECCV 2020)

DiffSfM (ICCV 2017)

RSSR (ICCV 2021)

CVR* (Ours)

CVR (Ours)

Ground-truth

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

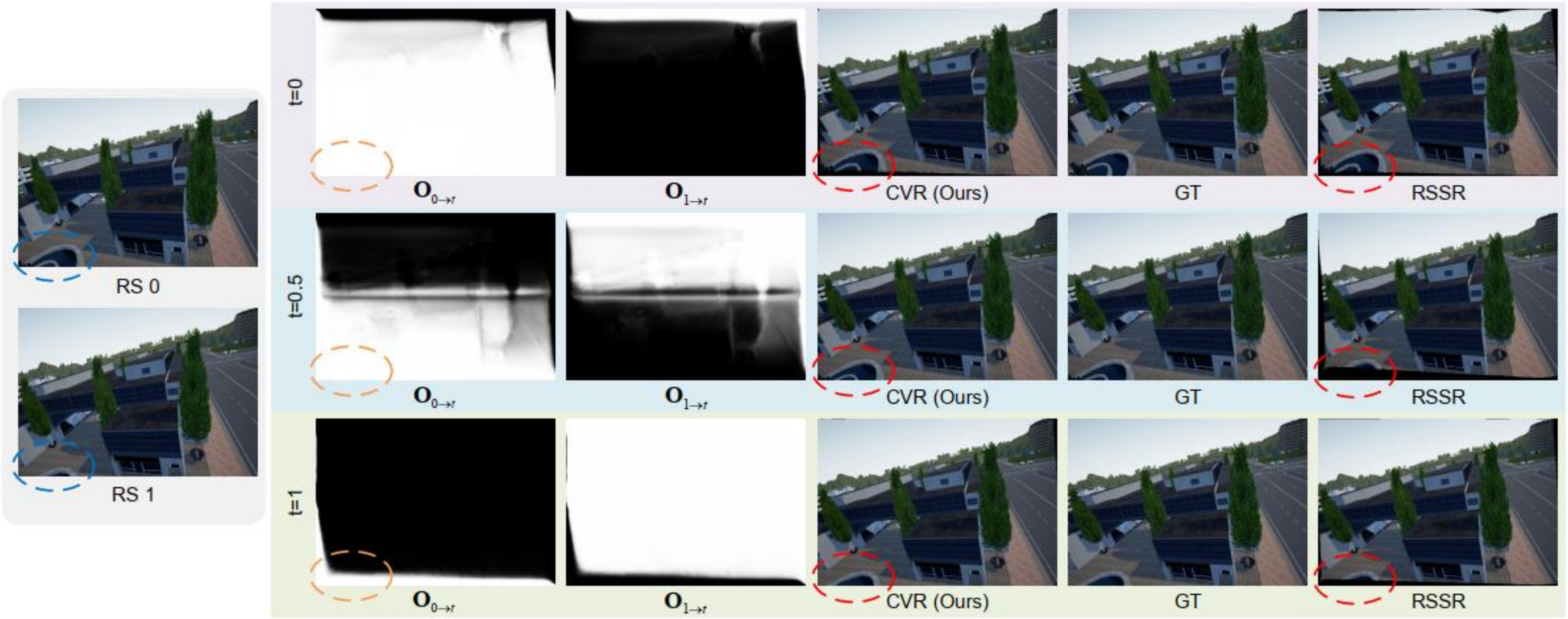
Experiments

- Effectiveness of our ABMF model

Input RS (Overlayed) | RSSR* | RSSR [9] | CVR* (Ours) | CVR (Ours) | Ground-truth

Experiments

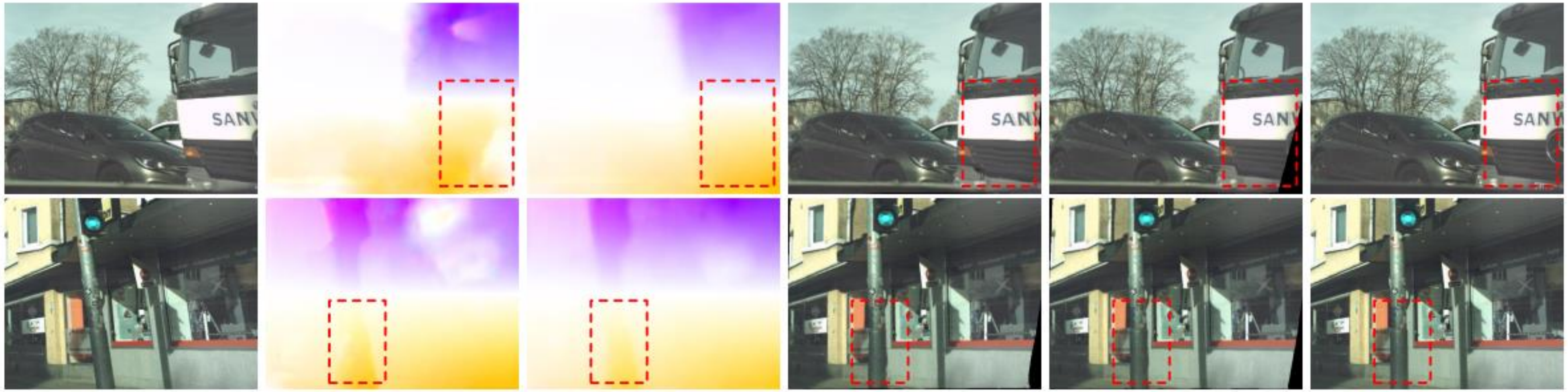
- Effectiveness of our occlusion reasoning layer



3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Effectiveness of our motion enhancement layer



Input RS 1 | Intermediate flow (RSSR) | Intermediate flow (CVR) | Estimated GS (RSSR) | Estimated GS (CVR) | Ground-truth

- ✓ The brighter a pixel, the bigger the motion enhancement.
- ✓ Our CVR effectively enhances ambiguous motion boundaries for more accurate contextual alignment.

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

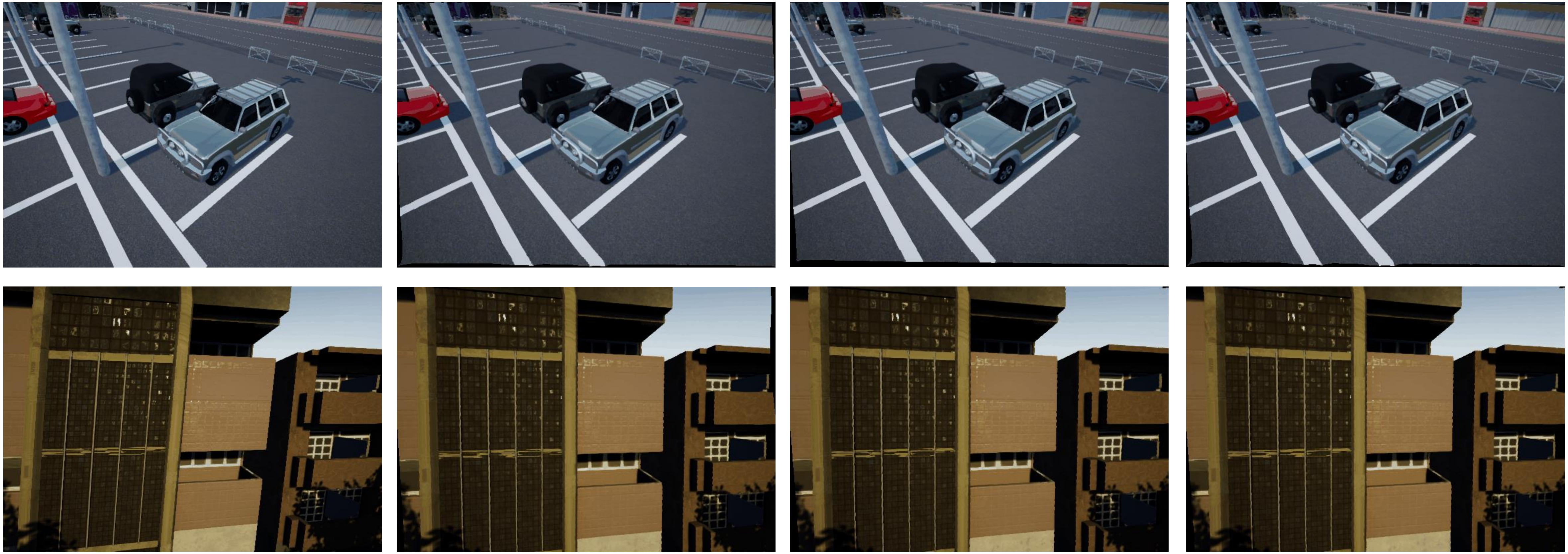
➢ Comparisons with video frame interpolation methods and cascaded methods

Input RS (Overlay) | BMBC | DAIN | Cascaded method | CVR (Ours) | Ground-truth

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Generating high-quality GS videos (Carla-RS dataset)



Input RS Frames | RSSR (ICCV 2021) | CVR* (Ours) | CVR (Ours)

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Generating high-quality GS videos (Fastec-RS dataset)

Input RS Frames | RSSR (ICCV 2021) | CVR* (Ours) | CVR (Ours)

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Generalizability on real data

Input RS Frames | RSSR (ICCV 2021) | CVR* (Ours) | CVR (Ours)

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

Experiments

- Generalizability on real data

Input RS Frames | RSSR (ICCV 2021) | CVR* (Ours) | CVR (Ours)

3.2. Context-Aware Video Reconstruction for Rolling Shutter Cameras

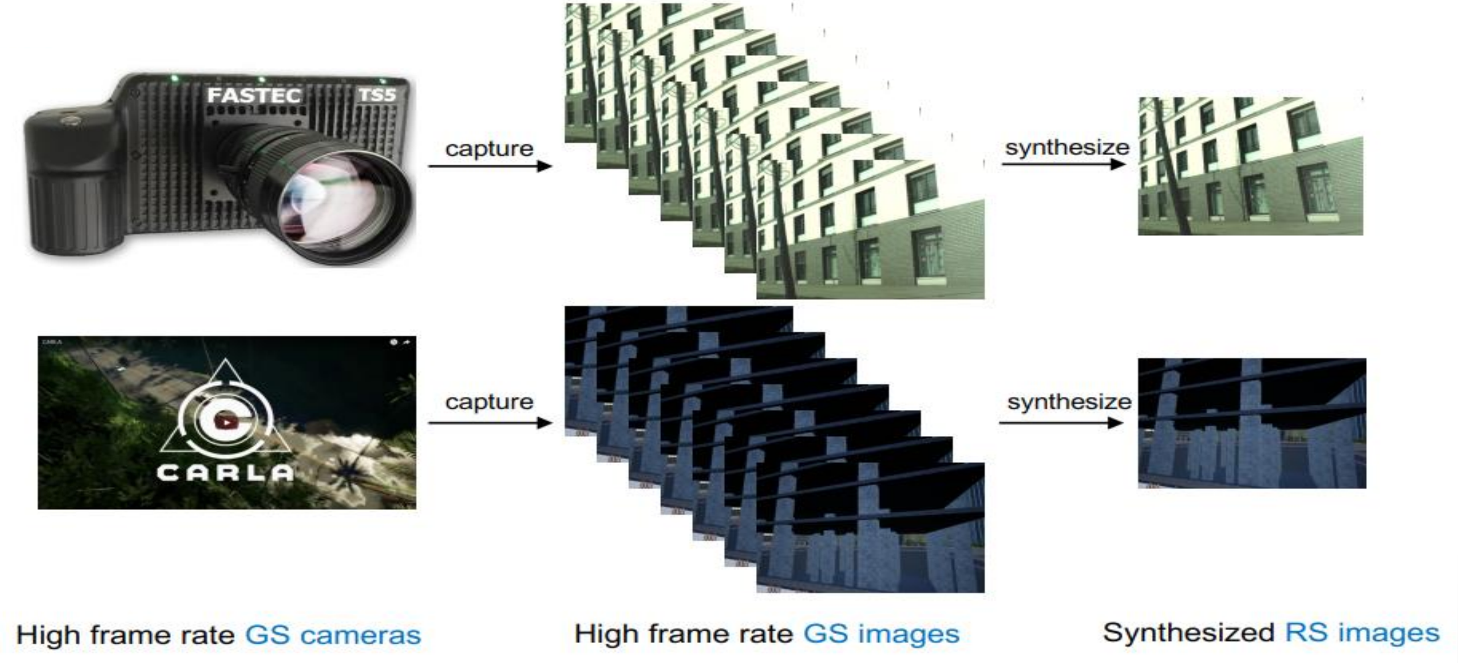
Conclusions

- We **re-define the RS temporal super-resolution problem** in the temporal dimension, which is beneficial for the temporally tractable joint RS correction and frame interpolation of RS video.
- We propose a simple yet effective **bilateral motion field approximation model**, which serves as a reliable initialization for GS frame refinement.
- We develop a stable and efficient **context-aware GS video reconstruction framework**, which can reason about complex occlusions, motion patterns specific to objects, and temporal abstractions.
- We demonstrate that the proposed method is more effective and compact than the SOTA approaches.

4. Public Datasets

Synthetic dataset:

- **Carla-RS**: It is generated from a virtual 3D environment using the Carla simulator, involving general 6-DoF camera motions. There are a training set of 210 sequences and a test set of 40 sequences, and each sequence consists of 10 consecutive frames. A total of **2500** RS images with a resolution of 640×448 pixels are included.
- **Fastec-RS**: It uses a high-speed GS camera mounted on the ground vehicle to collect high-FPS GS video sequences at 2400 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 **2584** RS image pairs with a resolution of 640×480 pixels.



✓ Note that they provide the GS ground-truth corresponding to the **first and middle scanlines** of the RS image.

Liu P, Cui Z, Larsson V, et al. Deep shutter unrolling network. CVPR 2020. (https://github.com/ethliup/DeepUnrollNet)

4. Public Datasets

Real-world dataset:

- **BS-RSC**: It is a realistic benchmark dataset, collected by a well-designed beam-splitter acquisition system in the dynamic urban environment. There are 50, 16, and 15 sequences for training (2500 image pairs), validation (800 image pairs), and testing (750 image pairs), respectively. The image resolution is 1024×768 pixels.

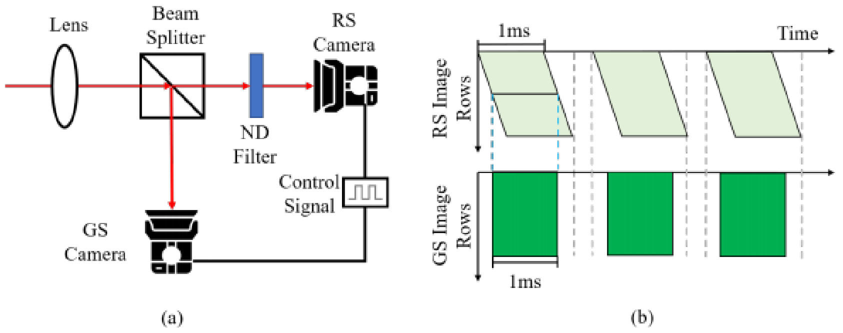


Figure 4. The designed beam-splitter acquisition system for real-world RSC dataset construction. **(a)** structure of the designed beam-splitter acquisition system. **(b)** exposure scheme of the GS and RS camera. The acquisition system can capture the GS frame at the intermediate exposure time of RS frame.

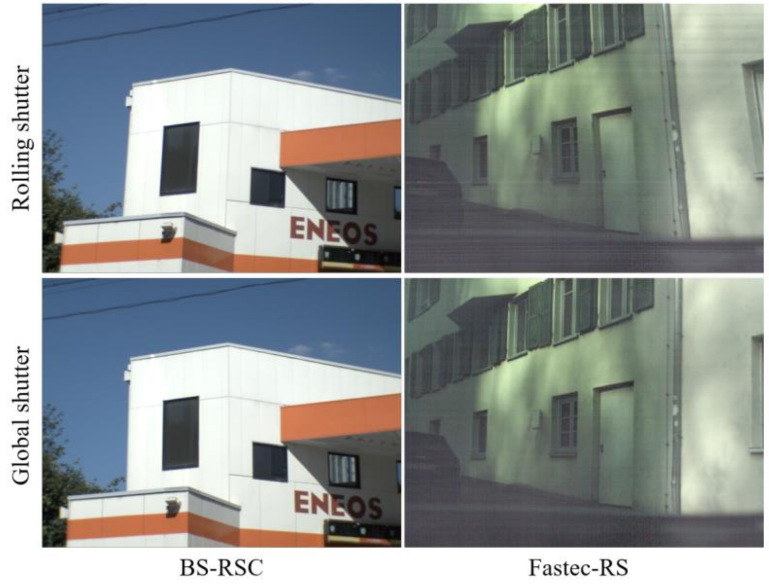


Figure 5. **Left**: The real world RS-GS example in the collected BS-RSC dataset. **Right**: The synthesized RS-GS example in the Fastec-RS dataset [20]. We see that our real RS frame is more natural, and there are much artifacts in the synthesized RS frames.

✓ Note that only the GS ground-truth corresponding to the **middle scanline** of the RS image is provided.

Cao M, Zhong Z, Wang J, et al. Learning adaptive warping for real-world rolling shutter correction. CVPR 2022. (https://github.com/ljzycmd/BSRSC)

4. Public Datasets

Real-world dataset:

- **BS-RSCD**: As a real dataset with egomotion 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., 4000 image pairs are recorded in total. The image resolution is 640×480 pixels.

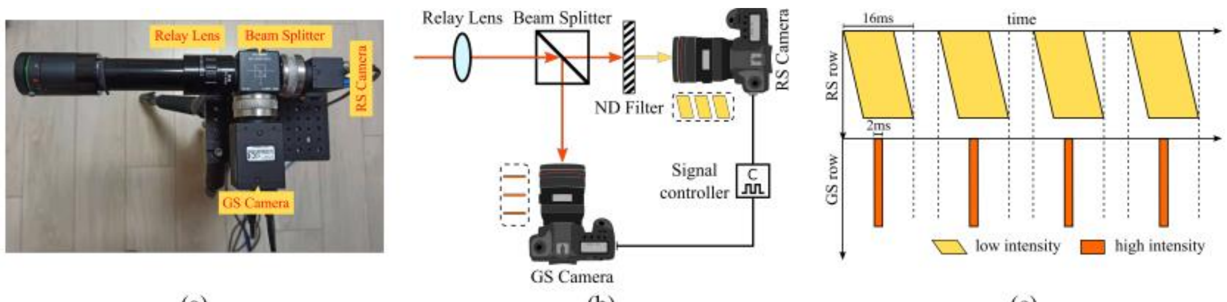


Figure 2: **Beam-splitter acquisition system.** (a) shows real system used to collect the dataset; (b) is system schematic diagram; (c) is exposure scheme of the system.

(a) Synthesized input (b) GT

(c) Realistic input (d) GT

✓ Note that only the GS ground-truth corresponding to the **middle scanline** of the RS image is provided.

Zhong Z, Zheng Y, Sato I. Towards rolling shutter correction and deblurring in dynamic scenes. CVPR 2021. (https://github.com/zzh-tech/RSCD)

Conclusion

- We introduce the rolling shutter correction method, mainly consisting of a dedistortion flow estimator and a GS image decoder.
- We introduce the RS temporal super-resolution method to reverse the rolling shutter imaging mechanism to generate a high-framerate and high-quality GS video.
- We introduce the RS dataset to enable efficient training of the above methods.

References

1. Fan B, Dai Y, Li H. Rolling shutter inversion: bring rolling shutter images to high framerate global shutter video. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022.
2. Fan B, Dai Y, Zhang Z, et al. Context-aware video reconstruction for rolling shutter cameras. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022: 17572-17582.
3. Fan B, Dai Y, He M. SUNet: symmetric undistortion network for rolling shutter correction. Proceedings of the IEEE International Conference on Computer Vision (**ICCV**), 2021: 4541-4550.
4. Fan B, Dai Y. Inverting a rolling shutter camera: bring rolling shutter images to high framerate global shutter video. Proceedings of the IEEE International Conference on Computer Vision (**ICCV**), 2021: 4228-4237.
5. Dai Y, Li H, Kneip L. Rolling shutter camera relative pose: generalized epipolar geometry. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2016: 4132-4140.
6. Cao M, Zhong Z, Wang J, et al. Learning adaptive warping for real-world rolling shutter correction. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022: 17785-17793.
7. Zhong Z, Zheng Y, Sato I. Towards rolling shutter correction and deblurring in dynamic scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2021: 9219-9228.
8. Zhong Z, Cao M, Sun X, et al. Bringing rolling shutter images alive with dual reversed distortion. European Conference on Computer Vision (**ECCV**), 2022: 233-249.
9. Liu P, Cui Z, Larsson V, et al. Deep shutter unrolling network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2020: 5941-5949.
10. Zhou X, Duan P, Ma Y, et al. EvUnroll: neuromorphic events based rolling shutter image correction. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022: 17775-17784.
11. Wang Z, Ji X, Huang J B, et al. Neural global shutter: learn to restore video from a rolling shutter camera with global reset feature. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022: 17794-17803.
12. Fan B, Dai Y, Wang K. Rolling-Shutter-stereo-aware motion estimation and image correction. Computer Vision and Image Understanding (**CVIU**), 2021, 213: 103296.
13. Fan B, Dai Y, Zhang Z, et al. Differential sfm and image correction for a rolling shutter stereo rig. Image and Vision Computing (**IVC**), 2022: 104492.

Thanks！

THANK YOU FOR YOUR ATTENTION

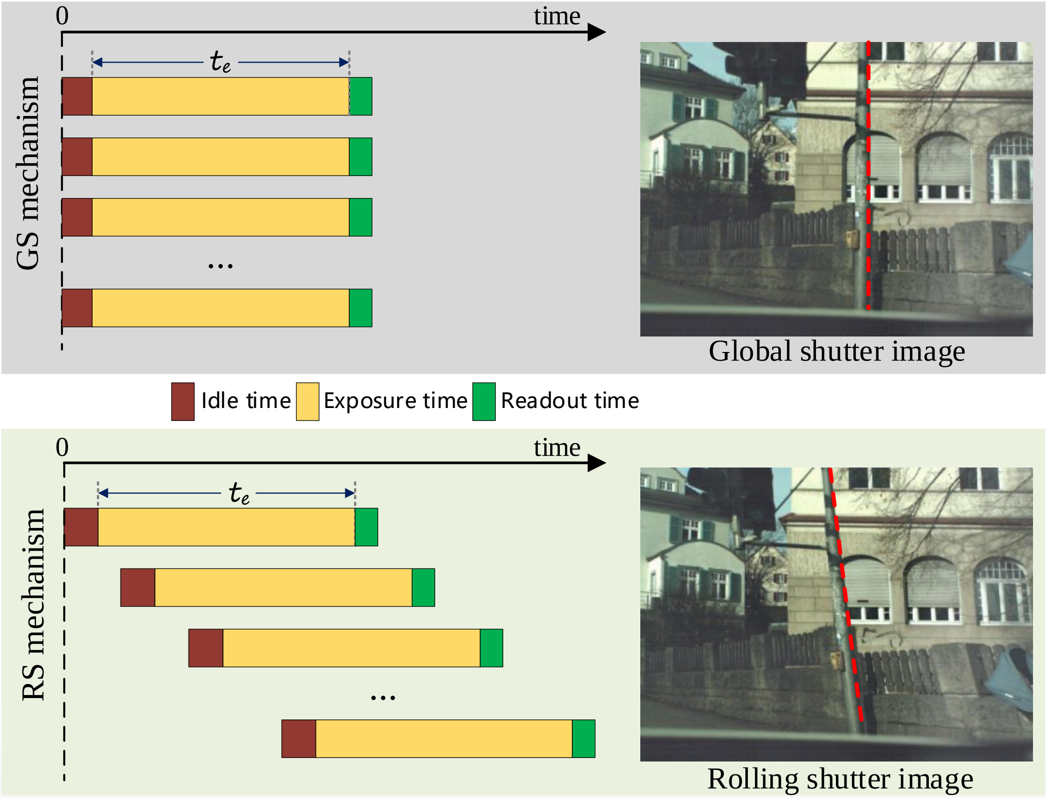
Email: daiyuchao@nwpu.edu.cn binfan@mail.nwpu.edu.cn

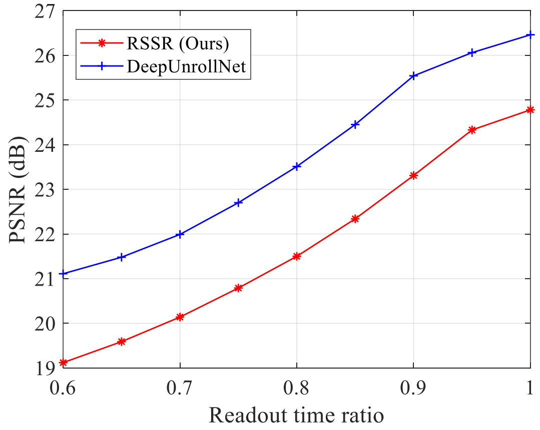
Further Direction and Discussion

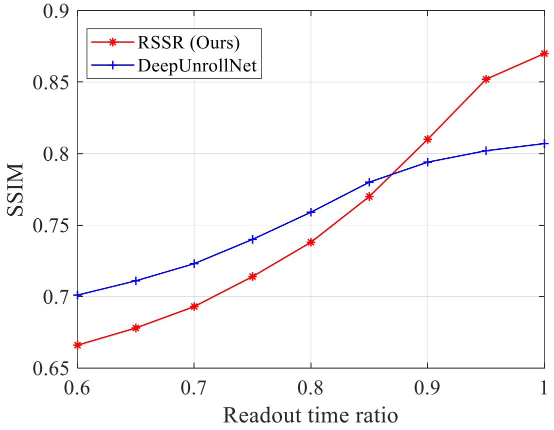
- **Lighter and more efficient network models.** Existing network architectures stack a large number of 2D convolutional modules to essentially achieve image-to-image translation, and thus are not yet capable of real-time GS image recovery, especially on low-power mobile devices. In addition, limited by the low resolution of the current training dataset, it will be a challenge to design lighter network models for high-resolution RS images (e.g. 4K 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.

Further Direction and Discussion

- **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.







Further Direction and Discussion

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

Related Papers

1. Zhou X, Duan P, Ma Y, et al. EvUnroll: Neuromorphic Events Based Rolling Shutter Image Correction. CVPR, 2022.
2. Wang Z, Ji X, Huang J B, et al. Neural Global Shutter: Learn to Restore Video from a Rolling Shutter Camera with Global Reset Feature. CVPR, 2022.
3. Zhong Z, Zheng Y, Sato I. Towards Rolling Shutter Correction and Deblurring in Dynamic Scenes. CVPR, 2021.
4. Mo J, Islam M J, Sattar J. IMU-Assisted Learning of Single-View Rolling Shutter Correction. Conference on Robot Learning, 2022.
5. Tourani S, Mittal S, Nagariya A, et al. Rolling Shutter and Motion Blur Removal for Depth Cameras. ICRA, 2016.

Further Direction and Discussion

- **Generate more realistic and multi-instant training datasets.** The current datasets either use a beam-splitter acquisition system to obtain ground truth GS images of real scenes, or simulate RS images by stitching row-by-row with high framerate GS videos. 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. To unleash the potential of deep learning methods, it is necessary to generate large-scale realistic RS datasets with more exposure instants, more diverse scenes and more dynamic objects.

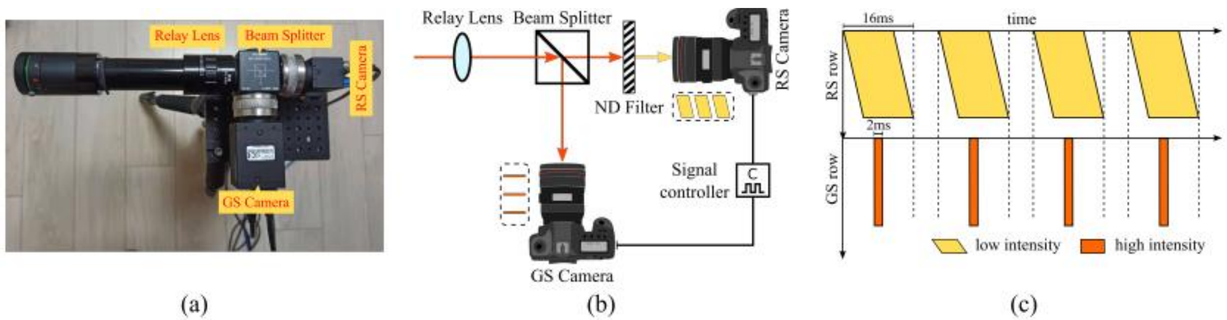


Figure 2: **Beam-splitter acquisition system.** (a) shows real system used to collect the dataset; (b) is system schematic diagram; (c) is exposure scheme of the system.

More realistic RS images