Rolling Shutter Image as an IMU

Prashanth Shyamala
University of Toronto

Motivation

- Camera pose and motion estimation is essential in many AR/VR and robotics applications. However, most existing techniques fail when rolling-shutter distortions introduce physical artifacts in the image.
- A common strategy is to use multiple frames and estimate the relative pose between them. Single-frame motion estimation is often avoided because the problem is ill-posed.
- Many methods assume a global-shutter camera model, but this assumption rarely holds for consumer devices, which predominantly use rolling shutter sensors.
- Rolling-shutter images encode temporal information across image rows, providing a rich cue that can be exploited to recover the underlying camera motion.

Related Work

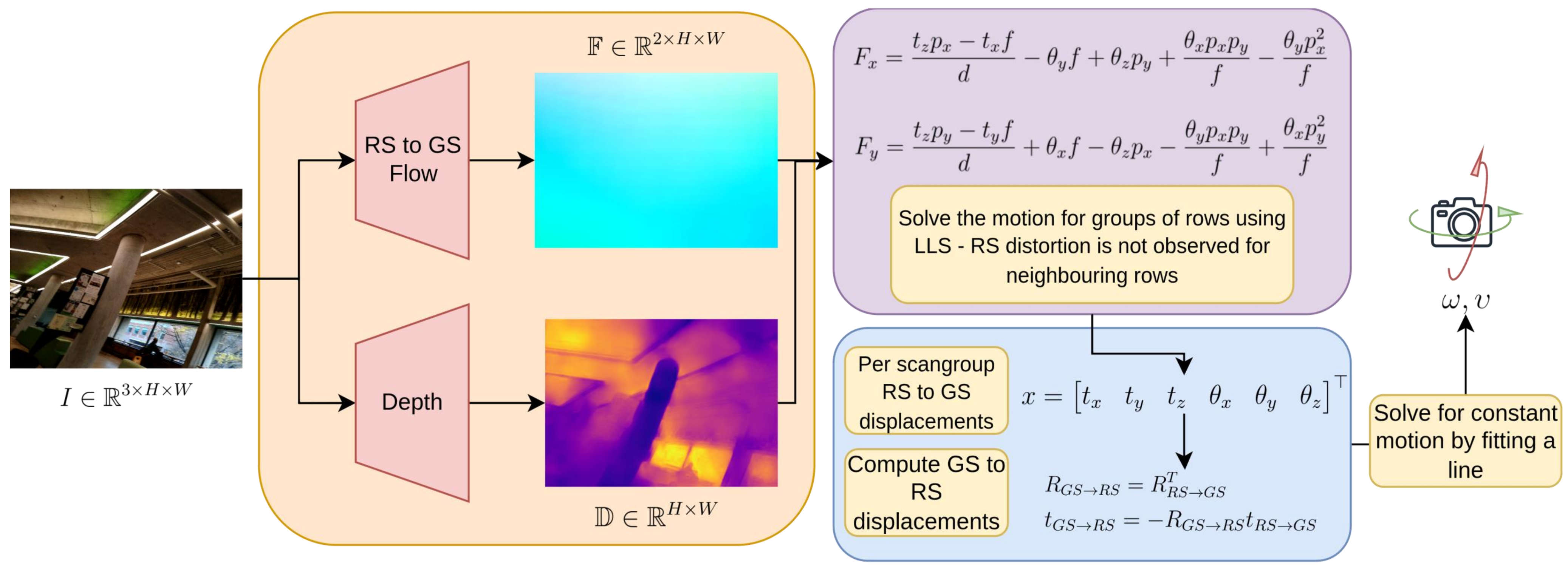
- **Image as an IMU [1] –** Uses image blur as a cue for motion. Predicts a motion field and depth from a single image, then estimates camera velocity.
- **R6P [2] - Rolling Shutter Absolute Pose Problem** - Solves for pose and motion using a polynomial solver on 3D–2D point correspondences under a rolling-shutter model.
- **A Generic Rolling Shutter Camera Model and its Application to Dynamic Pose Estimation [3]** - Estimates rotation and translation by minimizing reprojection error on 3D–2D correspondences. Models **non-uniform velocity** within a single frame.

References

[1] Chen, Jerred, and Ronald Clark. "Image as an imu: Estimating camera motion from a single motion-blurred image." arXiv preprint arXiv:2503.17358 (2025).
[2] C. Albl, Z. Kukelova and T. Pajdla, "R6P - Rolling shutter absolute pose problem," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015
[3] Magerand, Ludovic and Adrien Bartoli. "A Generic Rolling Shutter Camera Model and its Application to Dynamic Pose Estimation." (2010).
[4]Yang, Zhanglei, et al. "Single image rolling shutter removal with diffusion models." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 9. 2025.
[5]Emanuele Trucco and Alessandro Verri. Introductory Techniques for 3-D Computer Vision, pages 178–184. Prentice Hall PTR, USA, 1998.

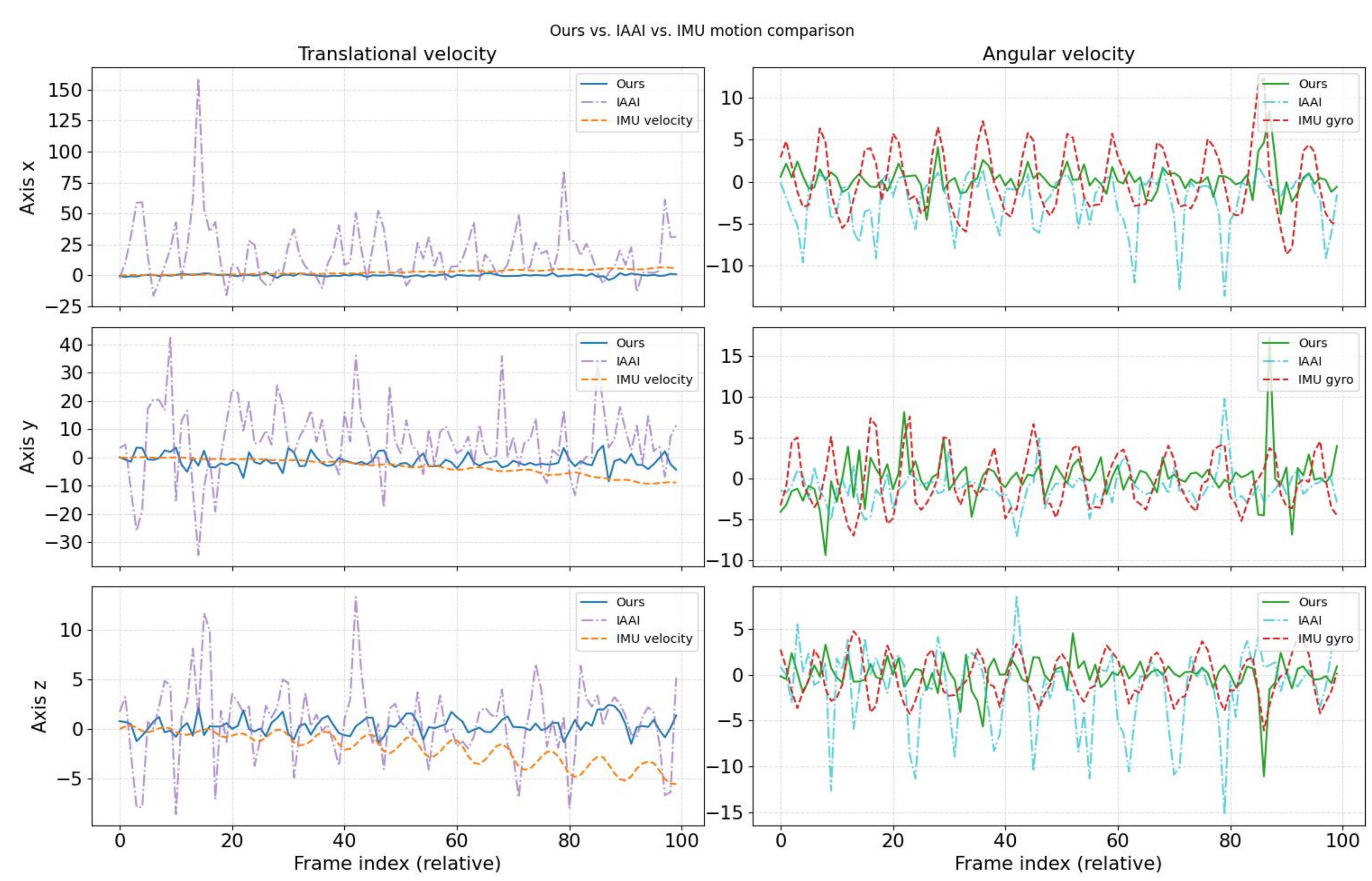
Methodology

I Extend "Image as an IMU" to Rolling-Shutter (RS) Images



- Predict RS→GS Flow Using a pretrained Diffusion Model [4]
- Predict RS-Aware Depth using the pretrained depth decoder from [1]
- Solve the motion-field equation using linear least squares (LLS) [5].
- Apply smoothing to remove high-frequency peaks in the estimated motion.
- Fit lines to the angular and translational displacement curves - The slopes correspond to the rotational and translational velocities.

Experimental Results



Method	Outdoor	Indoor	New College	Avg
Angular Velocities RMSE (rad/s)				
	$\omega_x/\omega_y/\omega_z$	$\omega_x/\omega_y/\omega_z$	$\omega_x/\omega_y/\omega_z$	$\omega_x/\omega_y/\omega_z$
IAAI	$3.95/\mathbf{3.15}/3.70$	$\mathbf{4.22}/\mathbf{3.71}/4.76$	$\mathbf{4.03}/3.52/4.23$	$4.07/\mathbf{3.46}/4.23$
Ours	$\mathbf{3.72}/3.95/\mathbf{2.75}$	$4.53/5.39/\mathbf{3.74}$	$4.15/\mathbf{2.70}/\mathbf{3.39}$	$\mathbf{4.13}/4.01/\mathbf{3.29}$
Translational Velocities RMSE (m/s)				
	$v_x/v_y/v_z$	$v_x/v_y/v_z$	$v_x/v_y/v_z$	$v_x/v_y/v_z$
IAAI	$23.03/127.73/4.60$	$56.64/151.12/4.22$	$42.42/188.19/21.37$	$40.70/155.68/10.73$
Ours	$\mathbf{3.39}/\mathbf{3.68}/\mathbf{2.90}$	$\mathbf{5.27}/\mathbf{7.81}/\mathbf{1.69}$	$\mathbf{3.52}/\mathbf{9.35}/\mathbf{18.37}$	$\mathbf{4.06}/\mathbf{6.95}/\mathbf{7.65}$

- The estimated angular velocities closely match the true rotation in both magnitude and direction, and the proposed method outperforms Image as an IMU (IAAI).
- The translational velocities match the peaks and troughs, but because the IMU outputs biased linear acceleration, integrating it over time causes a drift from the true ground-truth velocity.
- A key limitation lies in the RS→GS diffusion model, which assumes purely rotational motion. This introduces errors for scenes containing objects close to the camera.
- The depth model from IAAI is not RS-aware, leading to additional inaccuracies in depth and, consequently, motion estimation.
- Because the entire pipeline is differentiable, it can be trained end-to-end, which will improve motion accuracy and robustness.