

KFNet: Learning Temporal Camera Relocalization using Kalman Filtering

Lei Zhou¹, Zixin Luo¹, Tianwei Shen¹, Jiahui Zhang², Mingmin Zhen¹, Yao Yao¹, Tian Fang³, Long Quan¹

¹Hong Kong University of Science and Technology

²Tsinghua University

³Everest Innovation Technology

Motivation

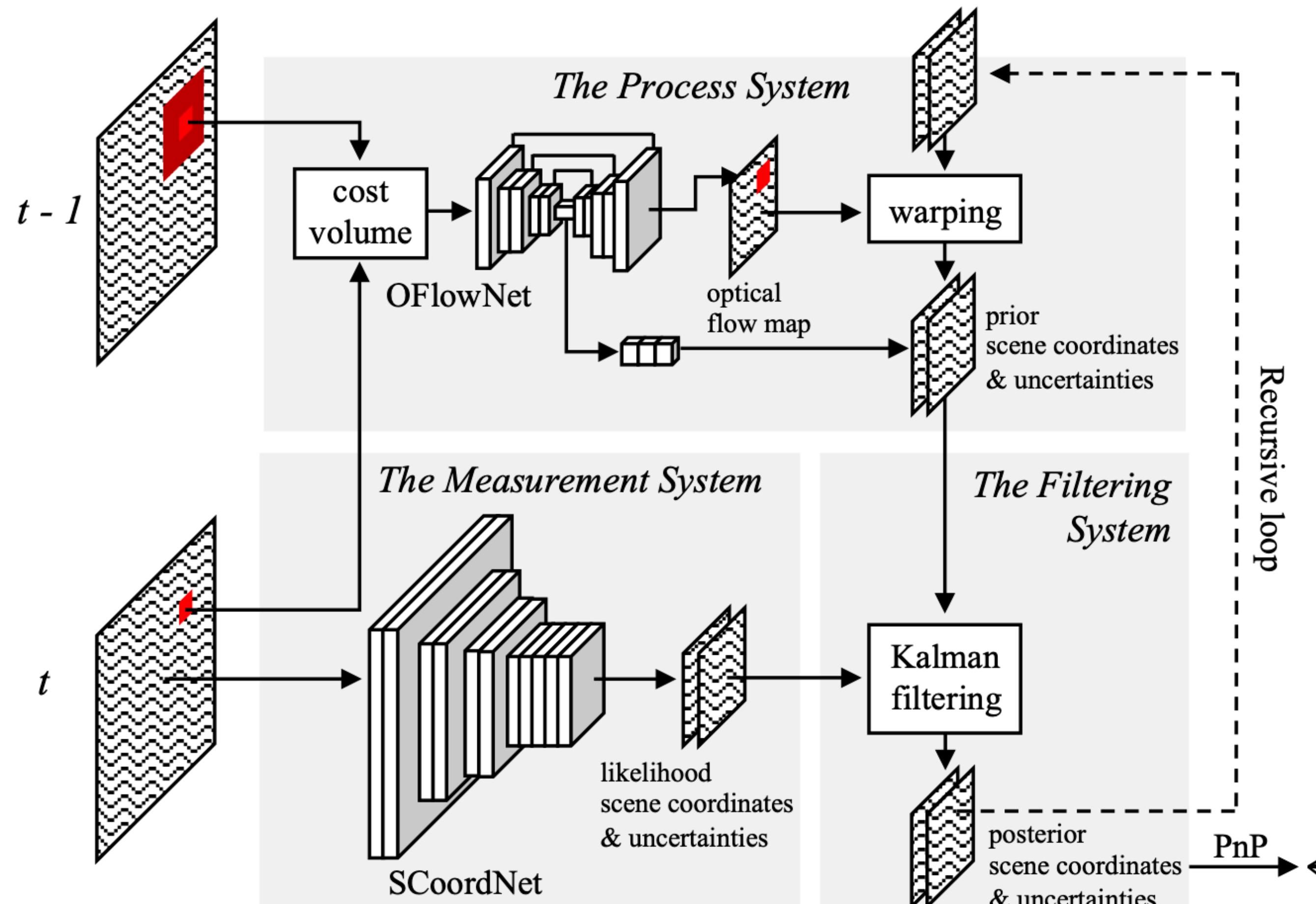
- Accurate image-based relocalization requires **2D-3D matches** and **projective geometry**. (Sattler, Torsten, et al.)
- While most works focus on one-shot relocalization, no efforts are made in temporal relocalization with 2D-3D matching in time domain.
- Temporal methods performs **even worse** than one-shot ones.

Contributions

- First to extend the **scene coordinate regression** problem to the time domain.
- Integrate the **Kalman filters** into a recurrent CNN network for pixel-level state inference over time-series images.
- Bridge the existing **performance gap** between temporal and one-shot relocalization approaches.

KFNet architecture

- Three sub-systems: the measurement, process and filtering systems.



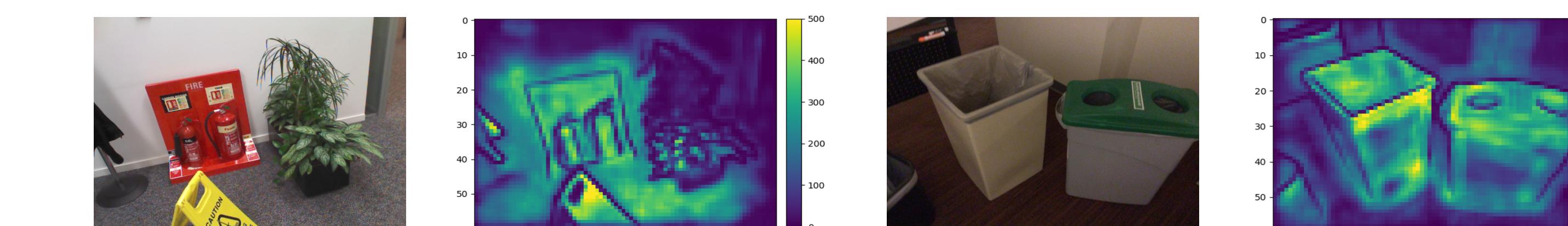
Bayesian formulation

The measurement system

- Generative model: image observations are generated from the underlying scene coordinate map, i.e., $P(I_t | y_t)$.
- Fully convolutional network: map I_t to z_t , then $P(z_t | y_t)$.
- Estimate Gaussian measurement noise for **likelihood loss**.

$$\mathcal{L}_{likelihood} = \sum_{i=1}^N \left(3 \log v_{(i)} + \frac{\|z_{(i)} - y_{(i)}\|_2^2}{2v_{(i)}^2} \right)$$

Measurement noise

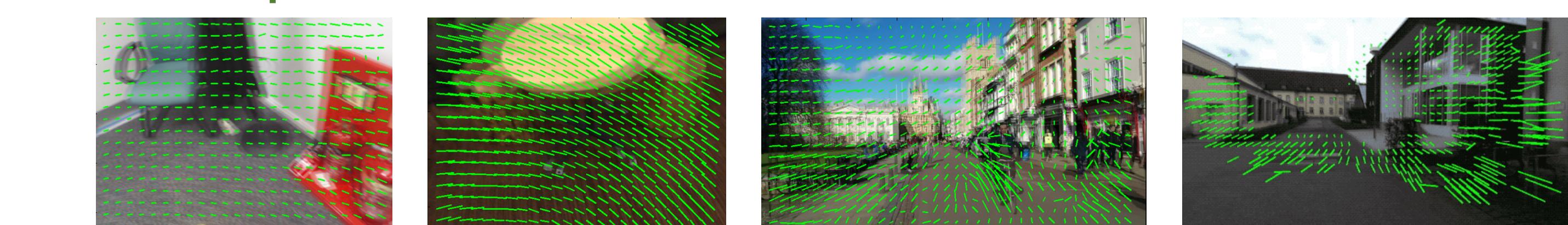


The process system

- Model the linear transition process by optical flow warping.
- Cost volume constructor + U-Net for flow estimation.
- Estimate Gaussian process noise for **prior loss**.

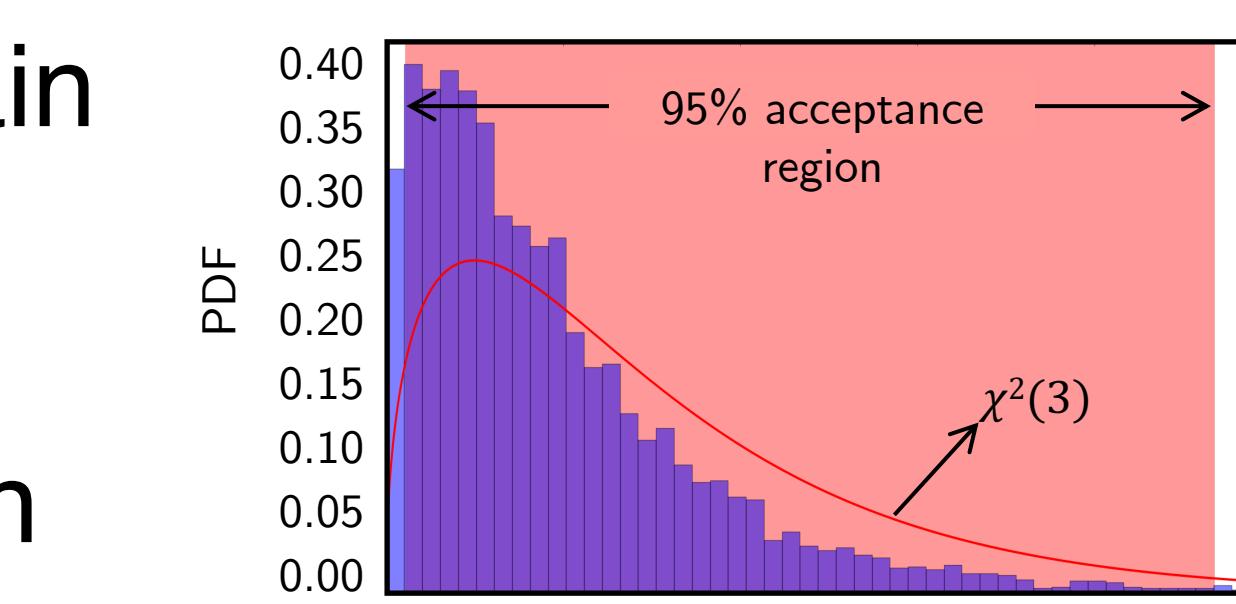
$$\mathcal{L}_{prior} = \sum_{i=1}^N \left(3 \log r_{(i)} + \frac{\|\hat{\theta}_{(i)}^- - y_{(i)}\|_2^2}{2r_{(i)}^2} \right)$$

Estimated optical flow fields



The filtering system

- Fusing both likelihood and prior estimations.
- Compute innovation and Kalman gain for **posterior loss**.
- NIS testing: negate potential outlier pixels outside the acceptance region



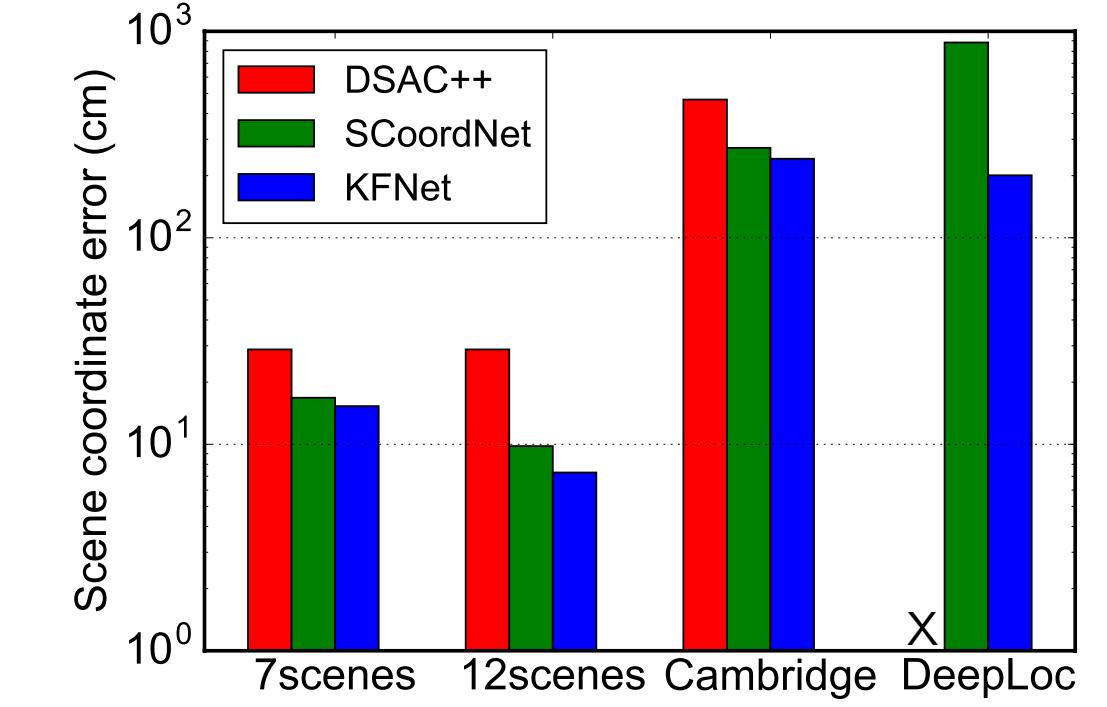
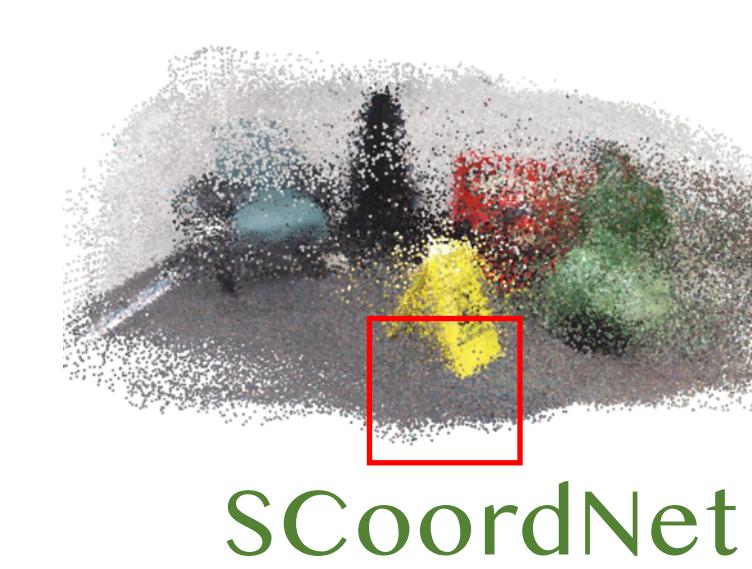
Total loss

$$\mathcal{L}_{posterior} = \sum_{i=1}^N \left(3 \log \sigma_{(i)} + \frac{\|\hat{\theta}_{(i)}^+ - y_{(i)}\|_2^2}{2\sigma_{(i)}^2} \right) \quad \mathcal{L}_{full} = \tau_1 \mathcal{L}_{likelihood} + \tau_2 \mathcal{L}_{prior} + \tau_3 \mathcal{L}_{posterior}$$

Results

The matching accuracy

- Right chart: mean error of scene coordinate predictions.
- KFNet** > **SCoordNet** > **DSAC++**

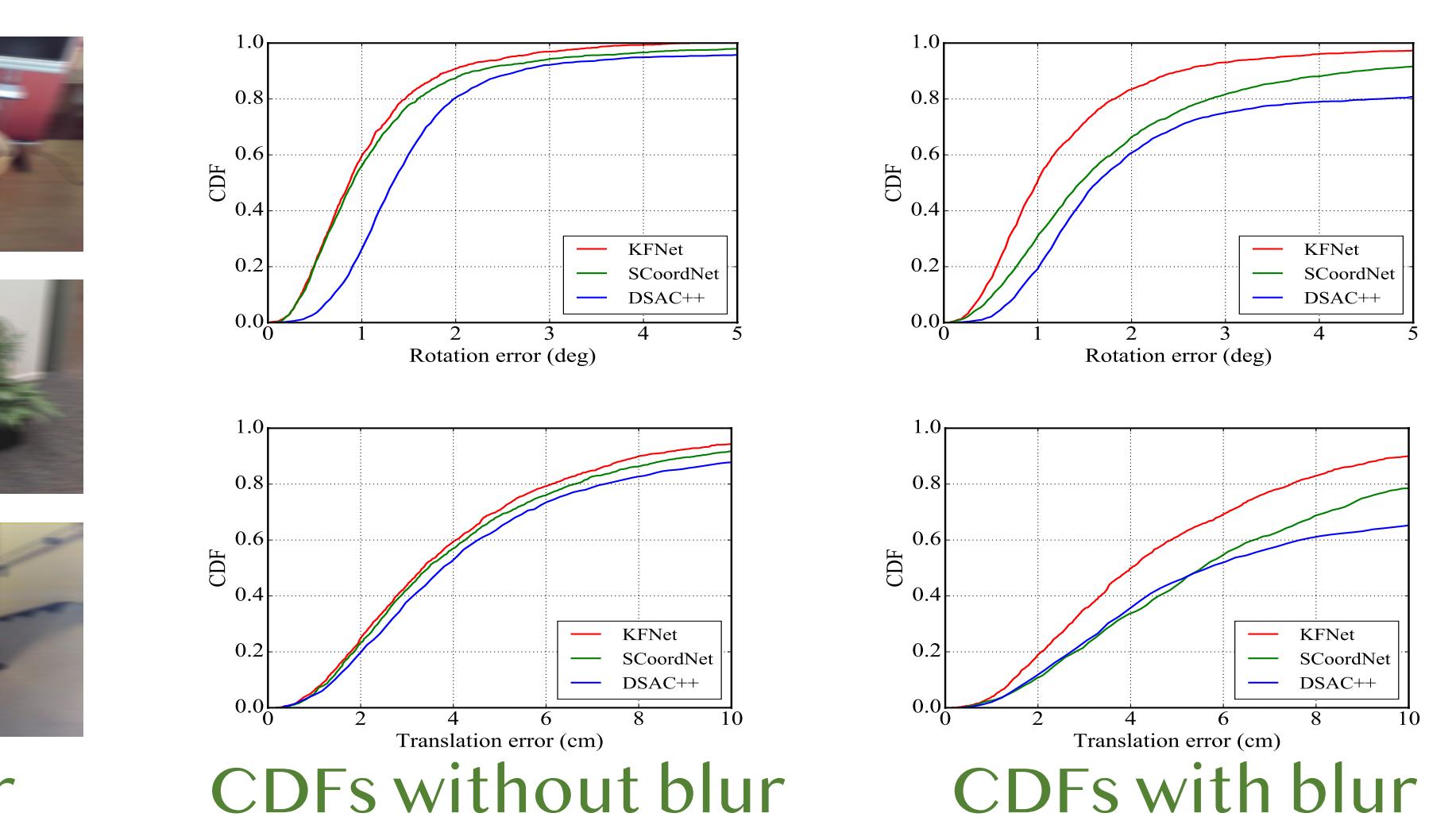
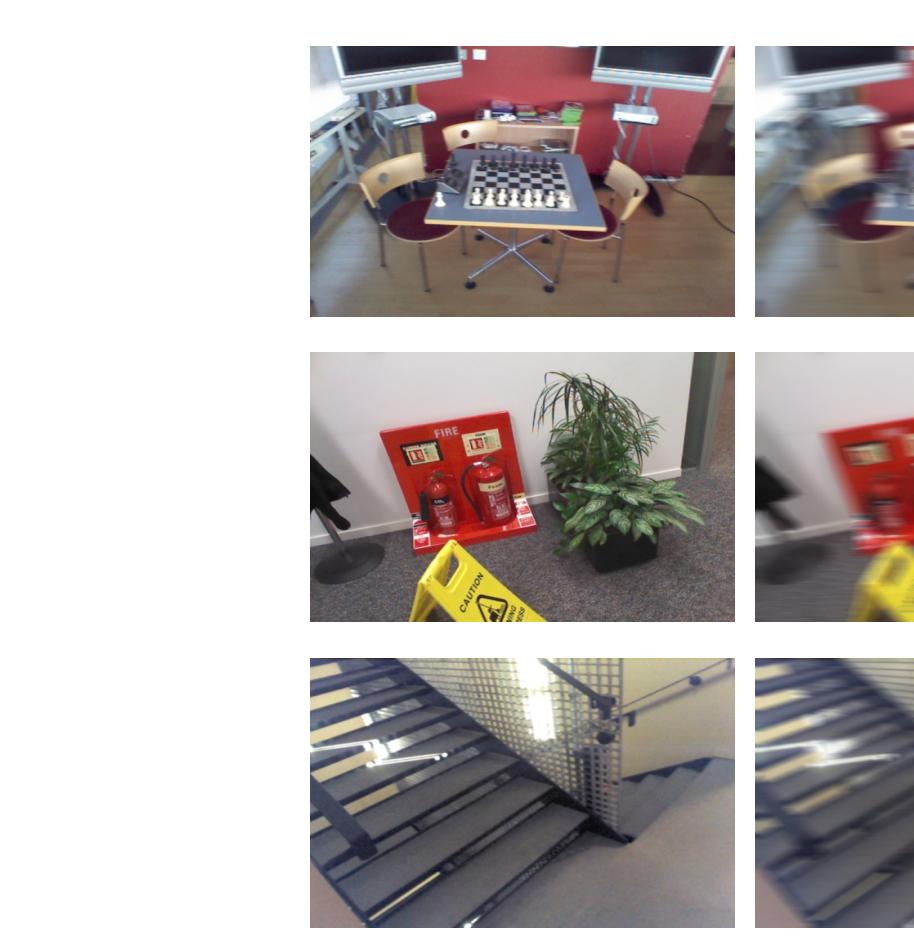


The relocalization accuracy

- Median translation and rotation error.

	One-shot relocalization					Temporal relocalization				
	MapNet	CamNet	AS	DSAC++	SCoordNet	VidLoc	LSTM-KF	VLocNet++	LSG	KFNet
7scenes	0.207m, 7.78°	0.040m, 1.69°	0.051m, 2.46°	0.036m, 1.10°	0.029m, 0.98°	0.246m,	0.424m, 11.00°	0.022m, 1.39°	0.190m, 7.47°	0.027m, 0.88°
Cambri dge	1.63m, 3.64°	-	0.29m, 0.63°	0.14m, 0.33°	0.13m, 0.32°	-	2.15m, 6.56°	-	-	0.13m, 0.30°
DeepLoc	-	-	-	-	-	0.083m, 0.45°	-	-	0.320m, 1.48°	-
										0.065m, 0.43°

- KFNet is more robust to motion blur.



Code release

- Code released at <https://github.com/zlthinker/KFNet>.
- Contact: <https://zlthinker.github.io/>.