Eye in the Sky: Drone-Based Object Tracking and 3D Localization

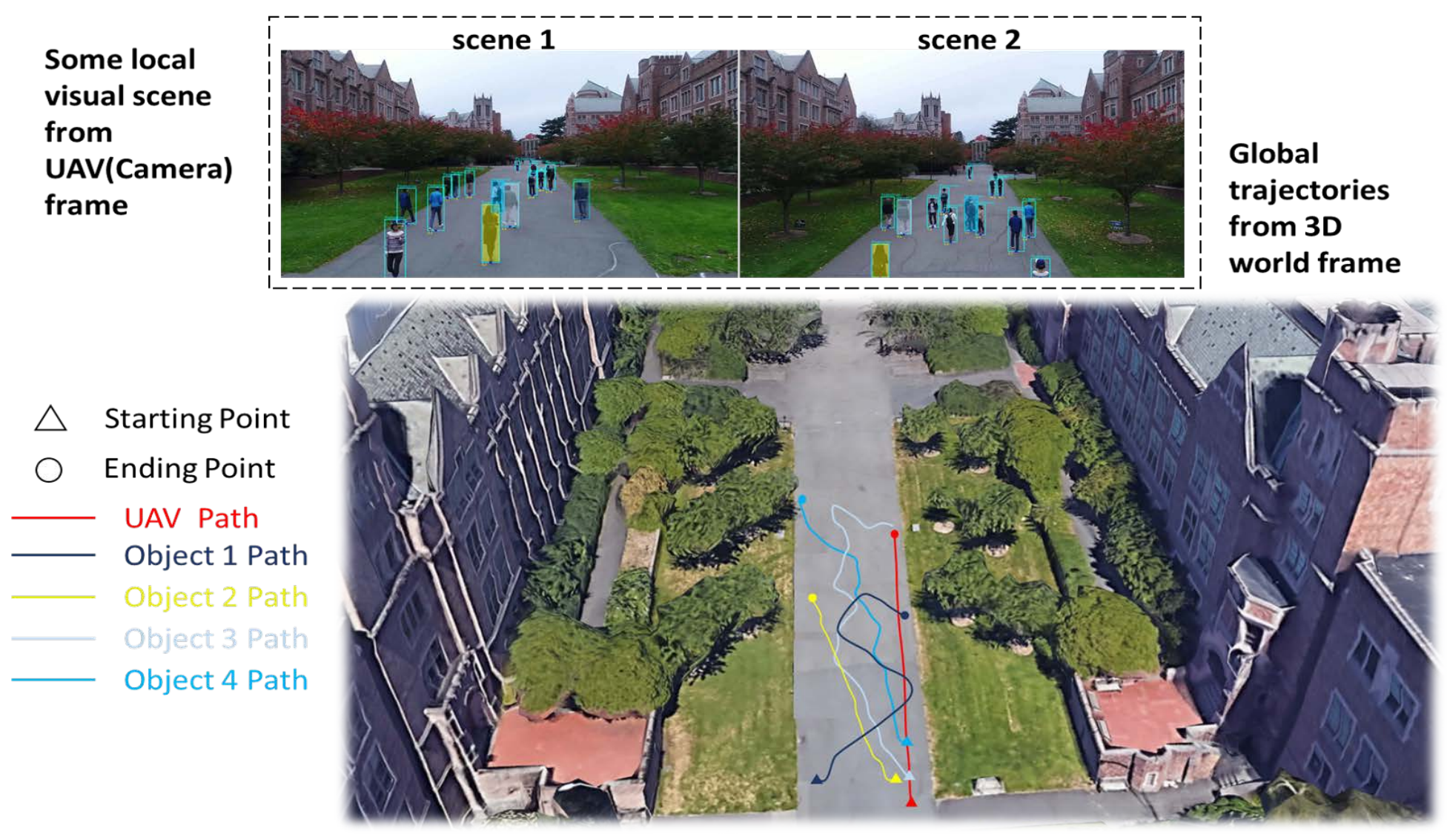
Haotian Zhang*, Gaoang Wang, Zhichao Lei, Jenq-Neng Hwang

Information Processing Lab, University of Washington, Seattle, WA, USA

Abstract

Drones, or general UAVs, equipped with a single camera have been widely deployed to a broad range of applications, such as aerial photography, fast goods delivery and most importantly, surveillance. Despite the great progress achieved in computer vision algorithms, these algorithms are not usually optimized for dealing with images or video sequences acquired by drones, due to challenges such as occlusion, fast camera motion and pose variation. In this paper, a drone-based multi-object tracking and 3D localization scheme is proposed based on the deep learning-based object detection. We first combine a multi-object tracking method called TrackletNet Tracker (TNT) which utilizes temporal and appearance information to track detected objects located on the ground for UAV applications. Then, we are also able to localize the tracked ground objects based on the group plane estimated from the Multi-View Stereo technique. The system deployed on the drone can not only detect and track the objects in a scene but can also localize their 3D coordinates in meters with respect to the drone camera. The experiments have proved our tracker can reliably handle most of the detected objects captured by drones and achieve favorable 3D localization performance when compared with the state-of-the-art methods.

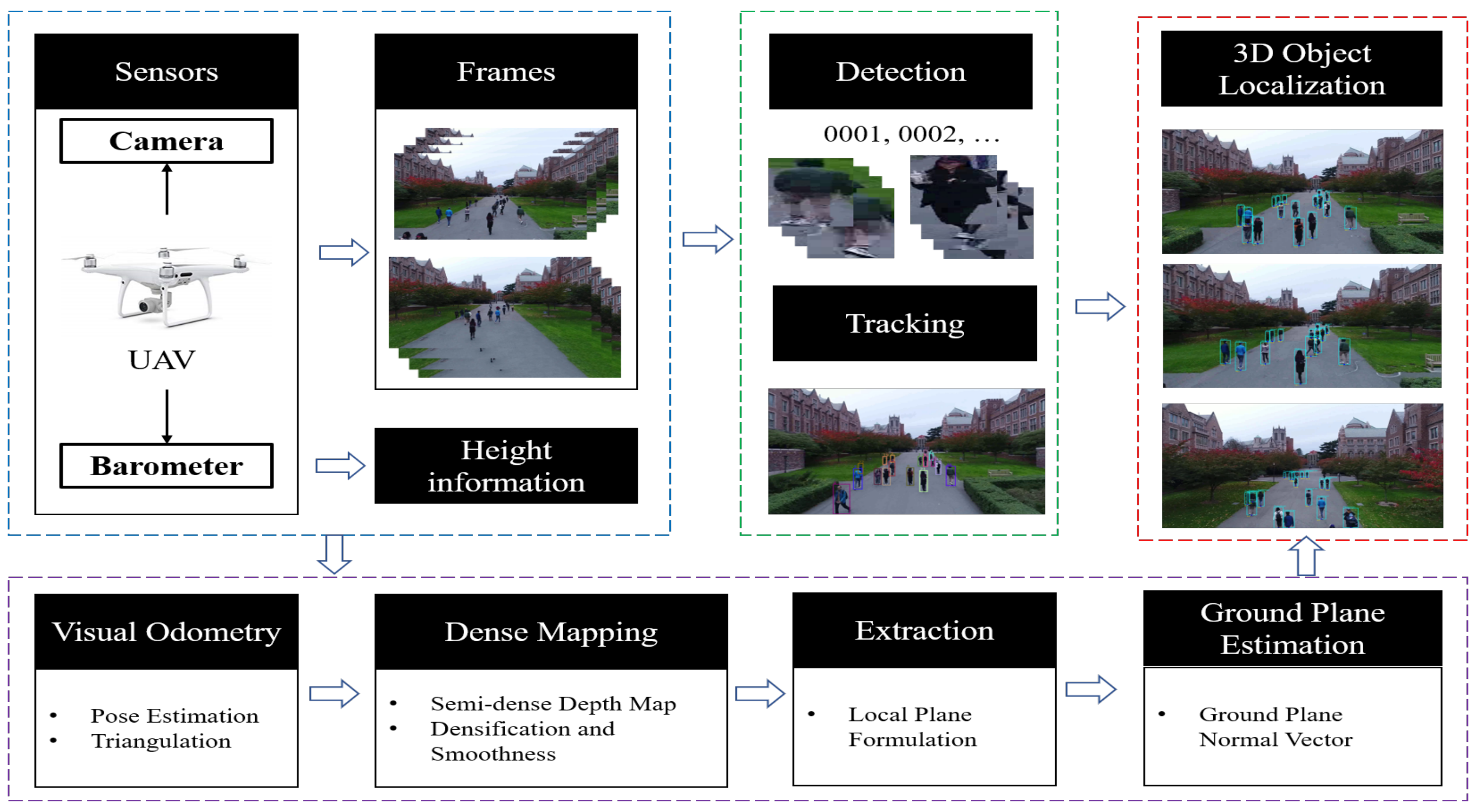
Introduction



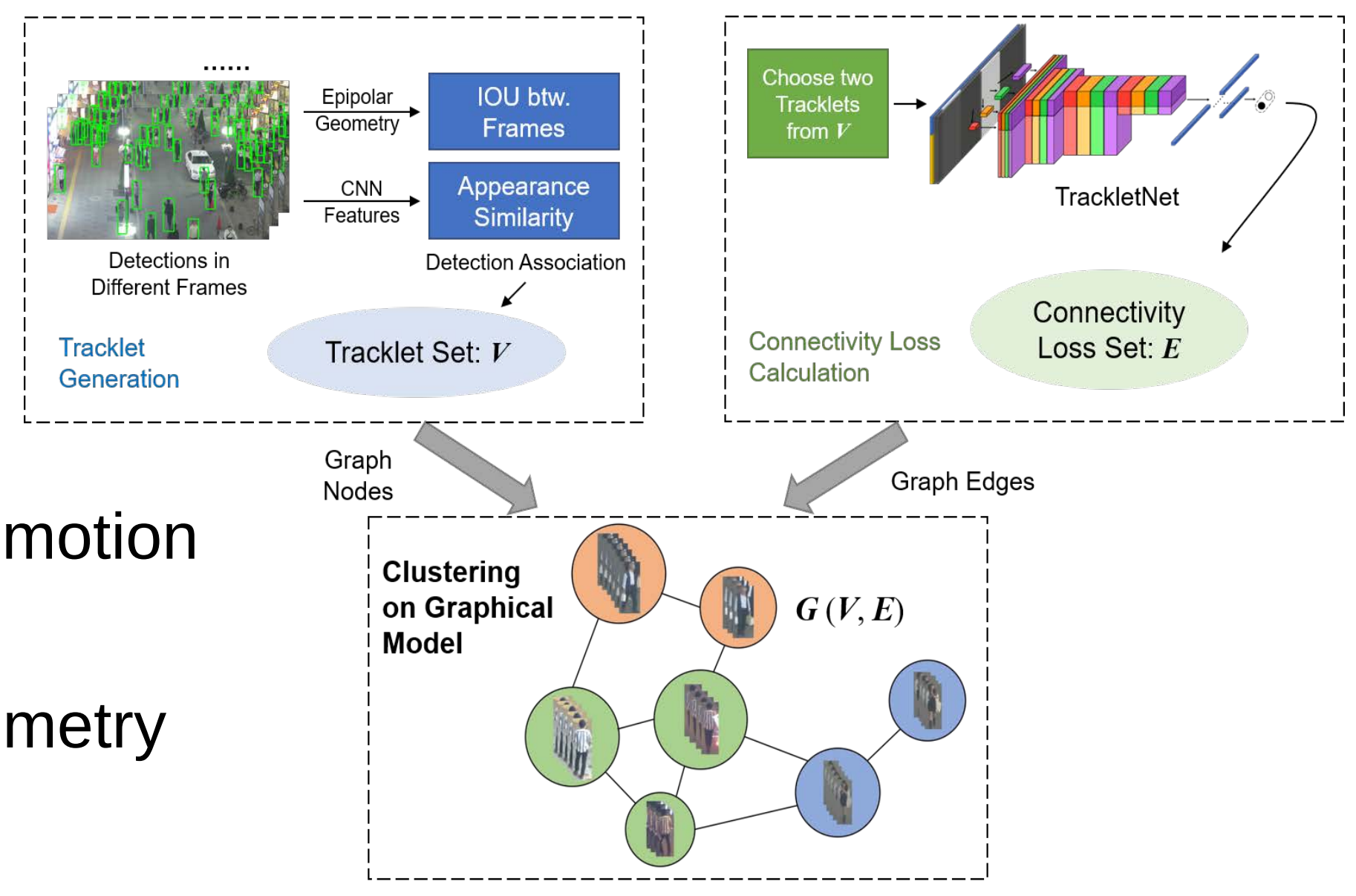
- ➢ *Deep-learning based object detection*
- ➢ *Multi-object tracking*
- ➢ *Visual Odometry and Ground Plane Estimation*
- ➢ *3D Object Localization*

Methods

System Flowchart*:*



Multi-object Tracking - TrackletNet



- *Bounding Box Association*

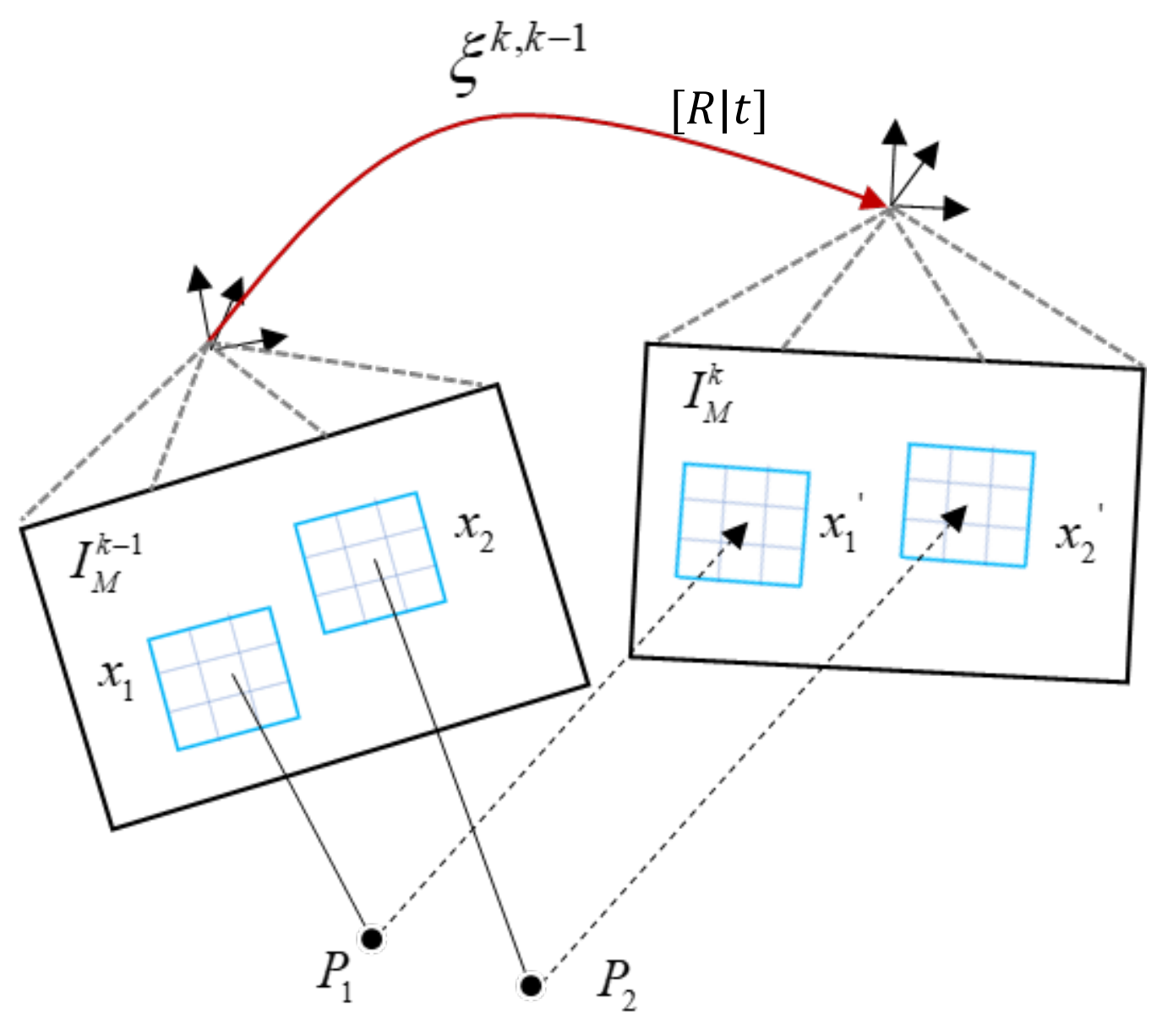
IOU + Appearance Similarity

- *Challenge*

Mis-association because of fast camera motion

- *Solution*

Bounding box prediction by epipolar geometry



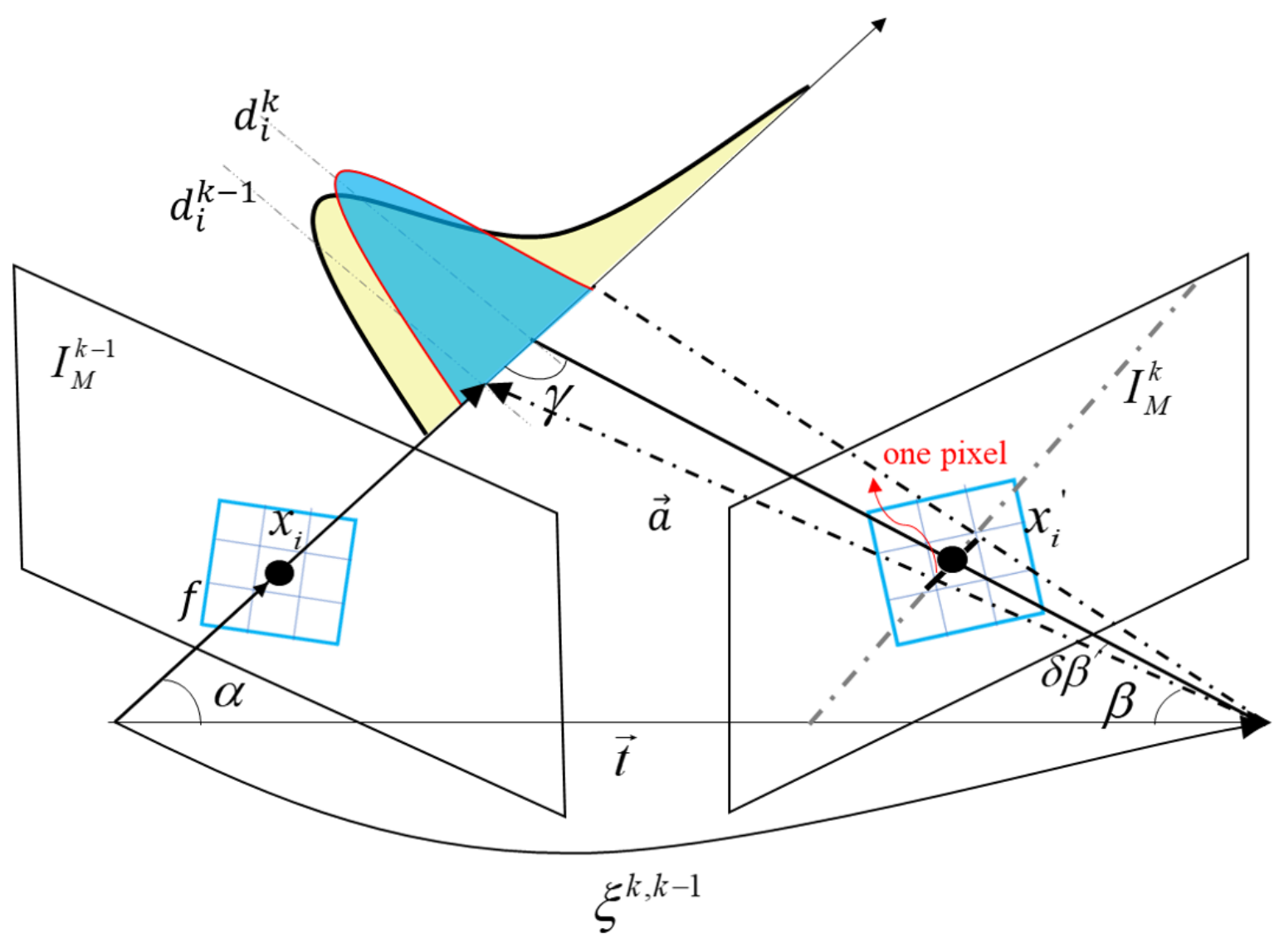
Semi-Direct Visual Odometry (SVO)

- *Minimizing the photometric error*

$$E(\xi) := \sum_{x \in \Omega_{D_M}} \|I_M(x) - I(\omega(x, D_M(x), \xi))\|_\delta$$

- *Extrinsic Camera parameters*: $[R|t]$

Ground Plane Estimation



- *Block Matching by Epipolar Constraint:*

$$S(x_i, x_i') = \frac{\sum_{m,n} x_i(m,n) x_i'(m,n)}{\sqrt{\sum_{m,n} x_i(m,n)^2 x_i'(m,n)^2}}$$

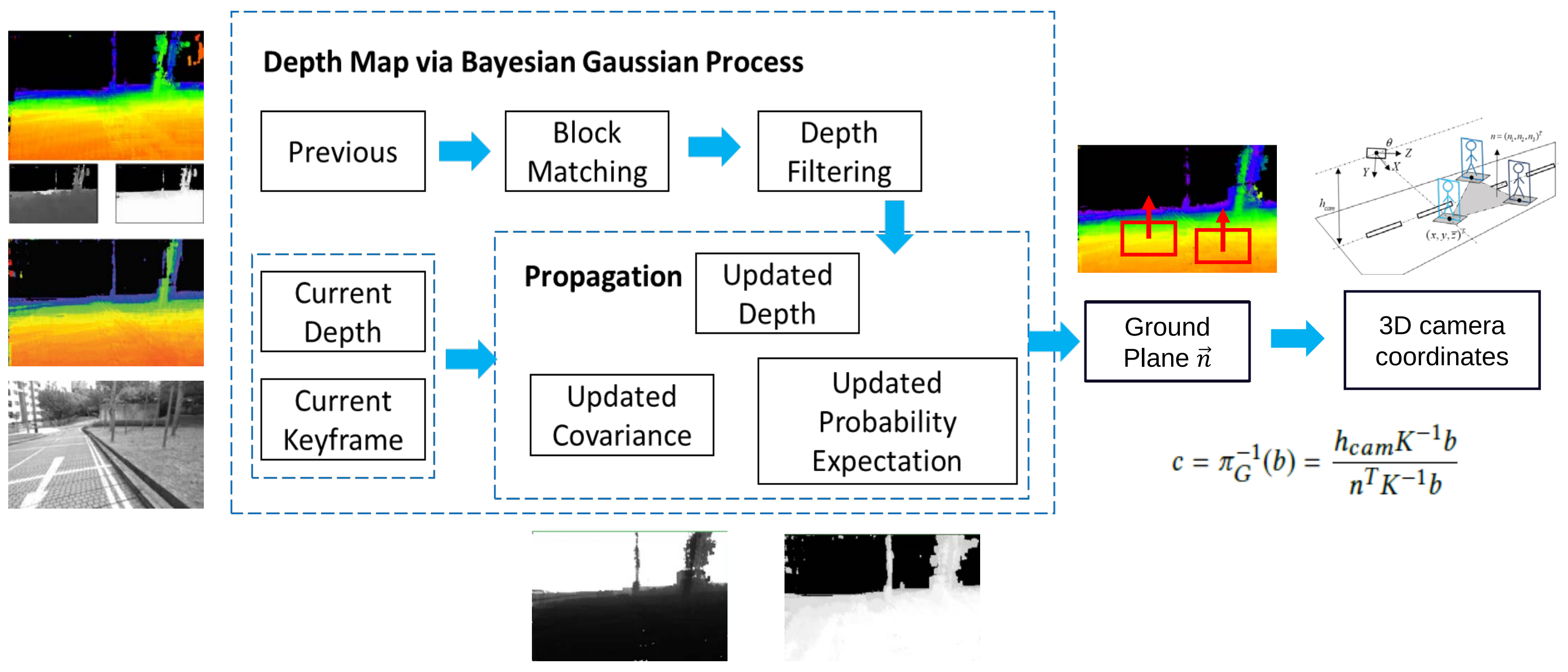
- *Bayesian Gaussian Depth Filter:*

inverse depth d $\quad p(d_i^k) \sim N(d_i^k | \mu_i, \sigma_i^2)$

$$p(\mu, \sigma^2 | d^1, \dots, d^N) \propto p(\mu, \sigma^2) \prod_k p(d^k | \mu, \sigma^2)$$

- *Multi-View Stereo (MVS) Method*

3D Localization



$$c = \pi_G^{-1}(b) = \frac{h_{cam} K^{-1} b}{n^T K^{-1} b}$$

Experiment Results

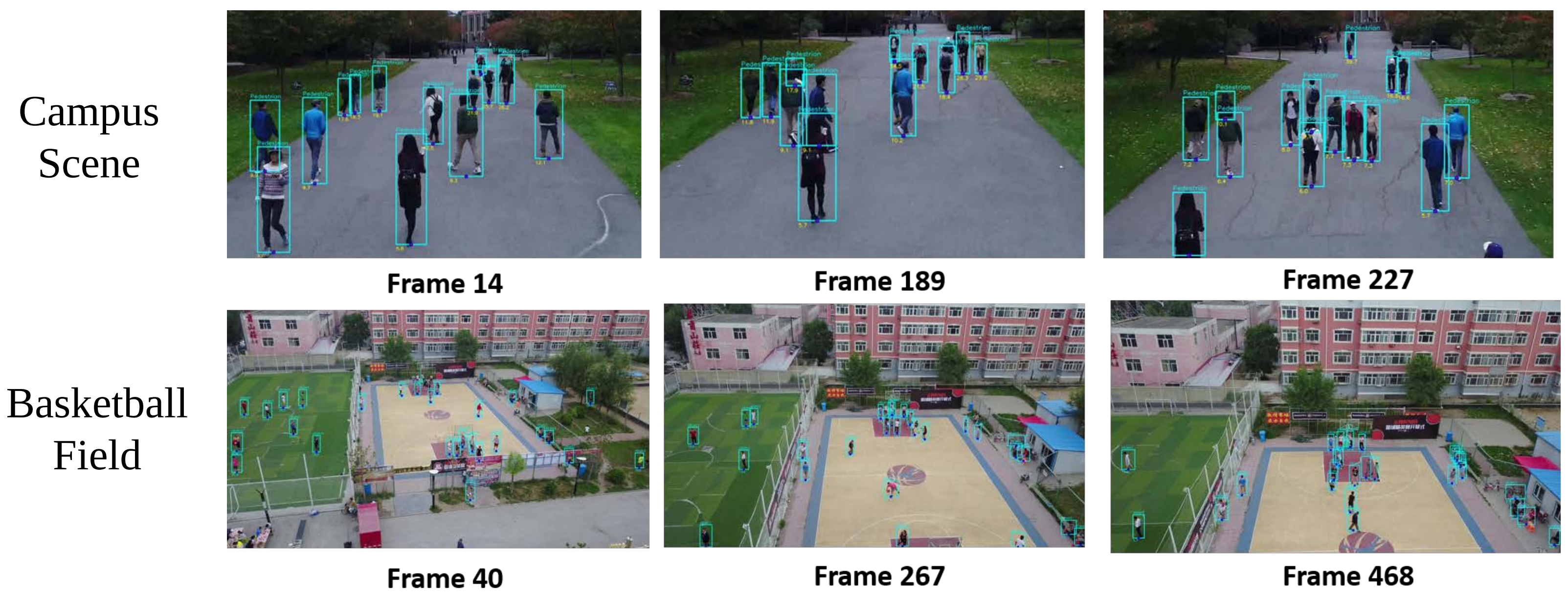


Table1: Tracking performance on the VisDrone2018-MOT test set compared to state-of-the-art. Best in **bold**, second best in blue.

Tracker	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	IDsw. ↓
V_IOU [5]	40.2	56.1	297	514	11,838	74,027	**265**
TrackCG [40]	42.6	58.0	323	395	14,722	68,060	779
GOG_EOC [25]	36.9	46.5	205	589	5,445	86,399	754
SCTrack [1]	35.8	45.1	211	550	7,298	85,623	798
Ctrack [41]	30.8	51.9	**369**	**375**	36,930	**62,819**	1,376
FRMOT [29]	33.1	50.8	254	463	21,736	74,953	1,043
GOG [25]	38.4	45.1	244	496	10,179	78,724	1,114
CMOT [2]	31.5	51.3	282	435	26,851	72,382	789
Ours	**48.6**	**58.1**	281	478	**5,349**	76,402	468

Table2: Mean localization error(standard deviation in parenthesis) in meters.

Approach	Scene	Overall (m)	<=10m	<=25m	>25m
Det+Flat_Ground_Asmp	Campus	3.84(±1.67)	4.05(±1.42)	4.76(±2.06)	N/A
	Grass land	3.96(±1.74)	2.41(±1.32)	3.98(±2.01)	N/A
	Basketball field	6.74(±3.15)	6.04(±2.78)	8.66(±3.18)	12.30(±3.84)
Det+Our_Ground_Est	Campus	2.22(±1.12)	2.04(±0.78)	2.61(±1.47)	N/A
	Grass land	2.27(±1.16)	1.15(±0.77)	1.98(±1.43)	N/A
	Basketball field	3.21(±1.84)	2.49(±1.66)	4.47(±2.12)	6.71(±2.33)
Det+Trk+Our_Ground_Est	Campus	0.49(±0.31)	0.47(±0.08)	1.21(±0.54)	N/A
	Grass land	0.78(±0.31)	0.21(±0.08)	0.94(±0.35)	N/A
	Basketball field	2.07(±1.46)	1.97(±1.22)	2.42(±1.74)	3.87(±1.95)

References

[1] Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Ling and Qinghua Hu, “Vision Meets Drones: A Challenge”, ECCV 2018.

[2] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. 2014. SVO: Fast semidirect monocular visual odometry. In 2014 IEEE international conference on robotics and automation (ICRA). IEEE, 15–22.

[3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2980–2988.

[4] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06), Vol. 1. IEEE, 519–528.