From a Bird's Eye View to See: Joint Camera and Subject Registration without the Camera Calibration

CVPR 2024

Zekun Qian[1], Ruize Han[#2,3], Wei Feng[1], Song Wang[4]

[1]College of Intelligence and Computing, Tianjin University

[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

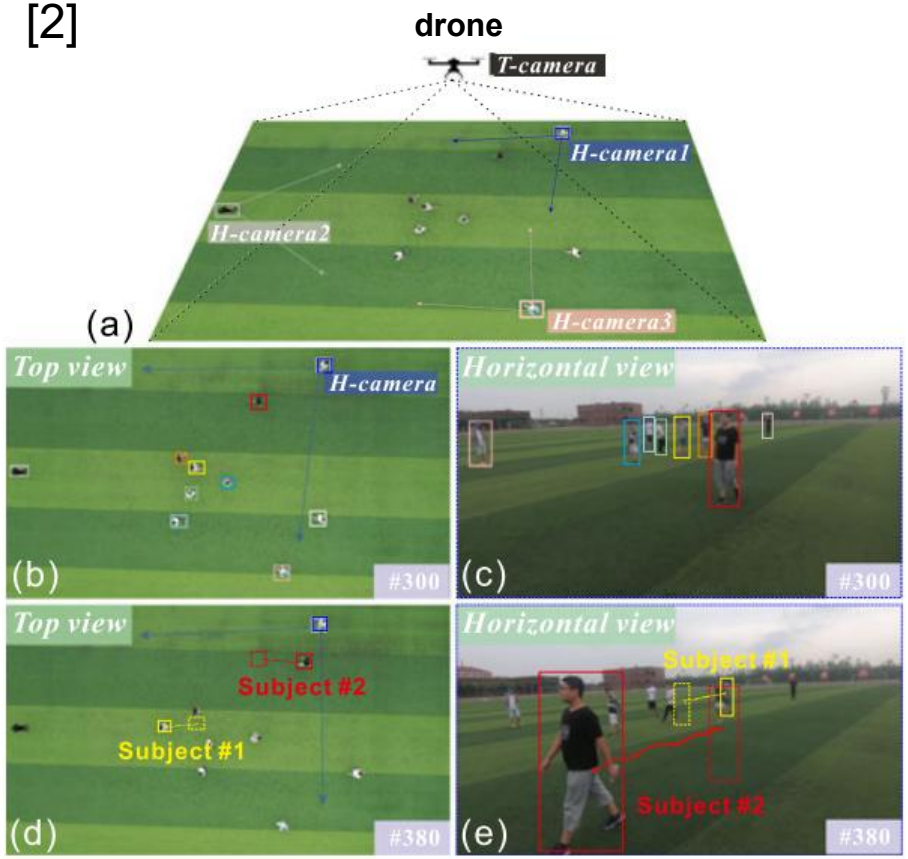
[3]City University of Hong Kong [4]University of South Carolina

Motivation

[1]

calibration

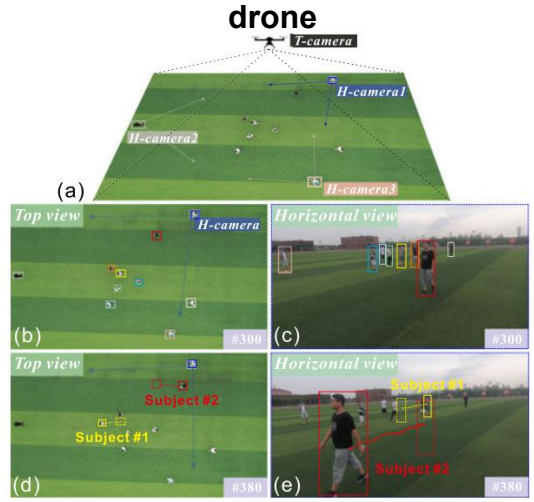
[2]



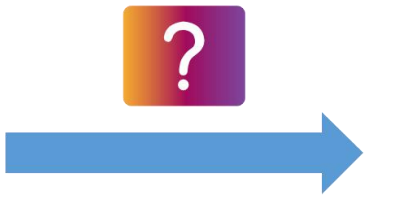
[1] Chavdarova, Tatjana, et al.. Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection. In CVPR, 2018.
[2] Han, Ruize, et al. Multiple human association and tracking from egocentric and complementary top views. IEEE TPAMI, 2021.

Motivation

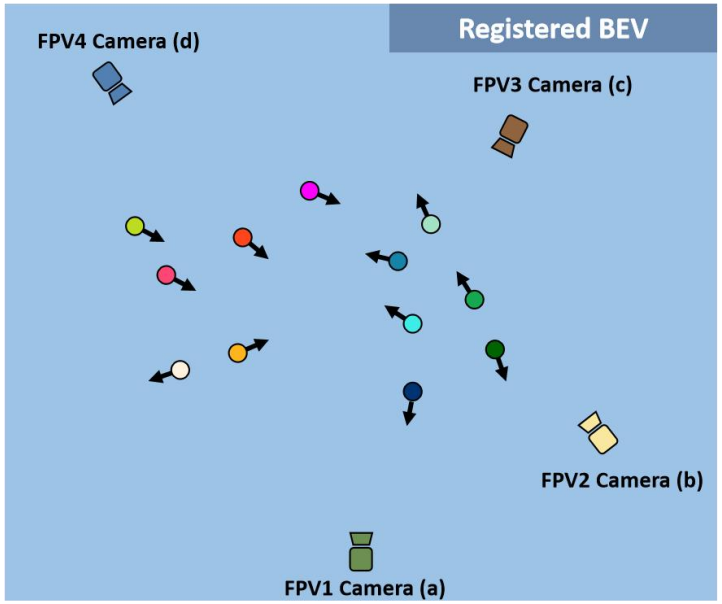
⇩ No camera calibration



⇩ No real BEV

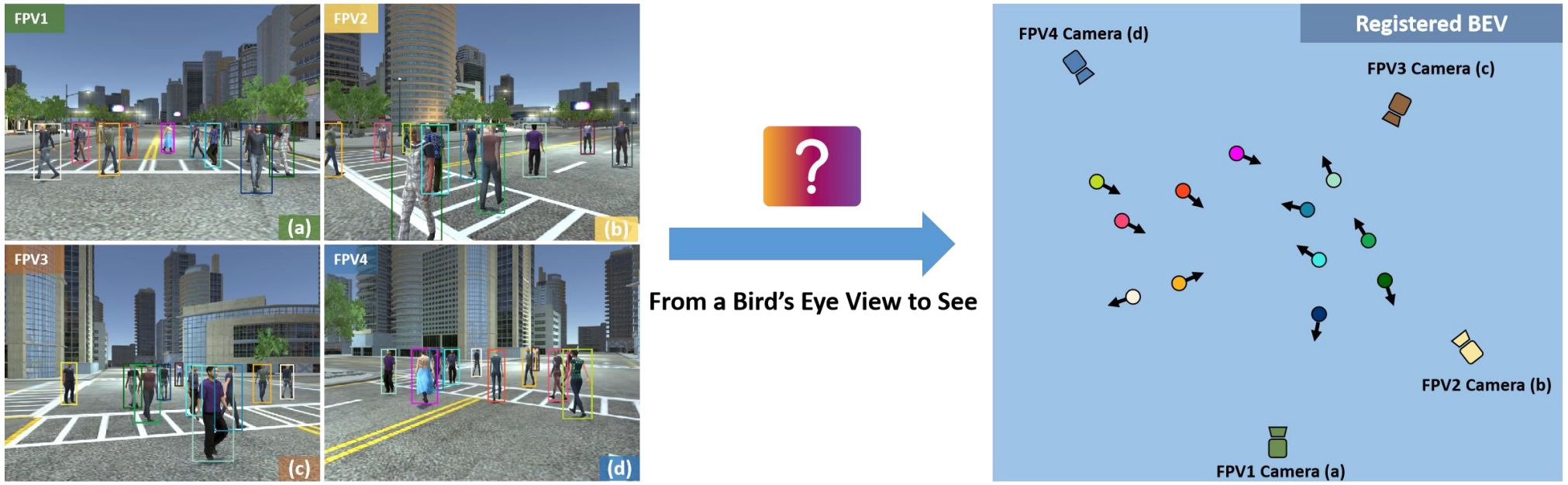


From a Bird's Eye View to See

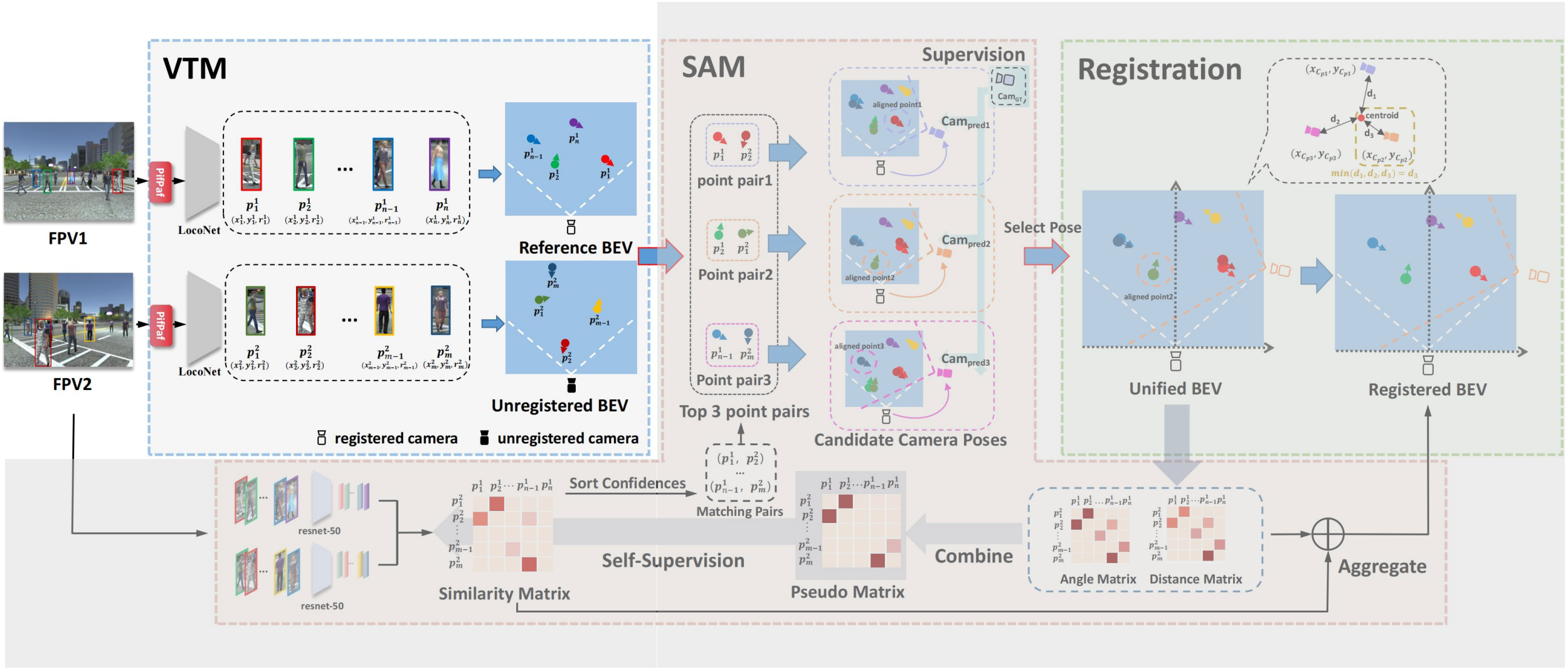


Challenges

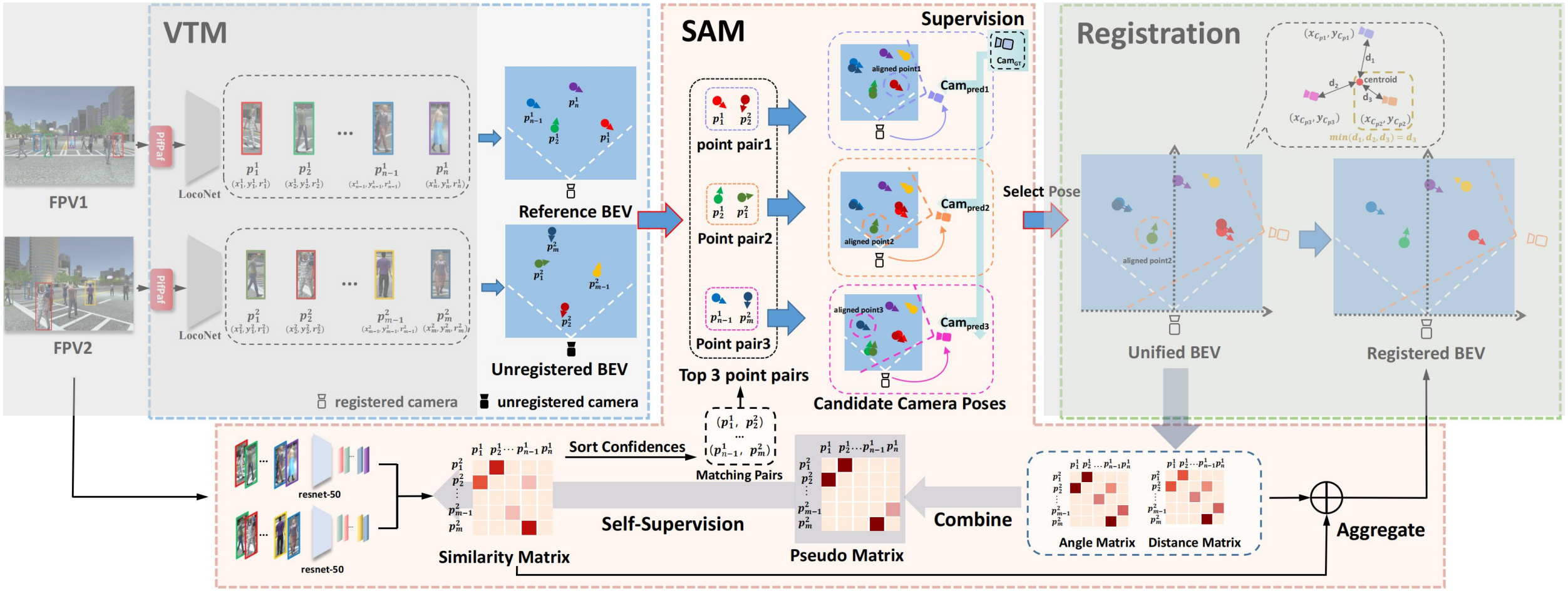
- Getting the camera registration without camera calibration
- Generating the registration results without the real BEV



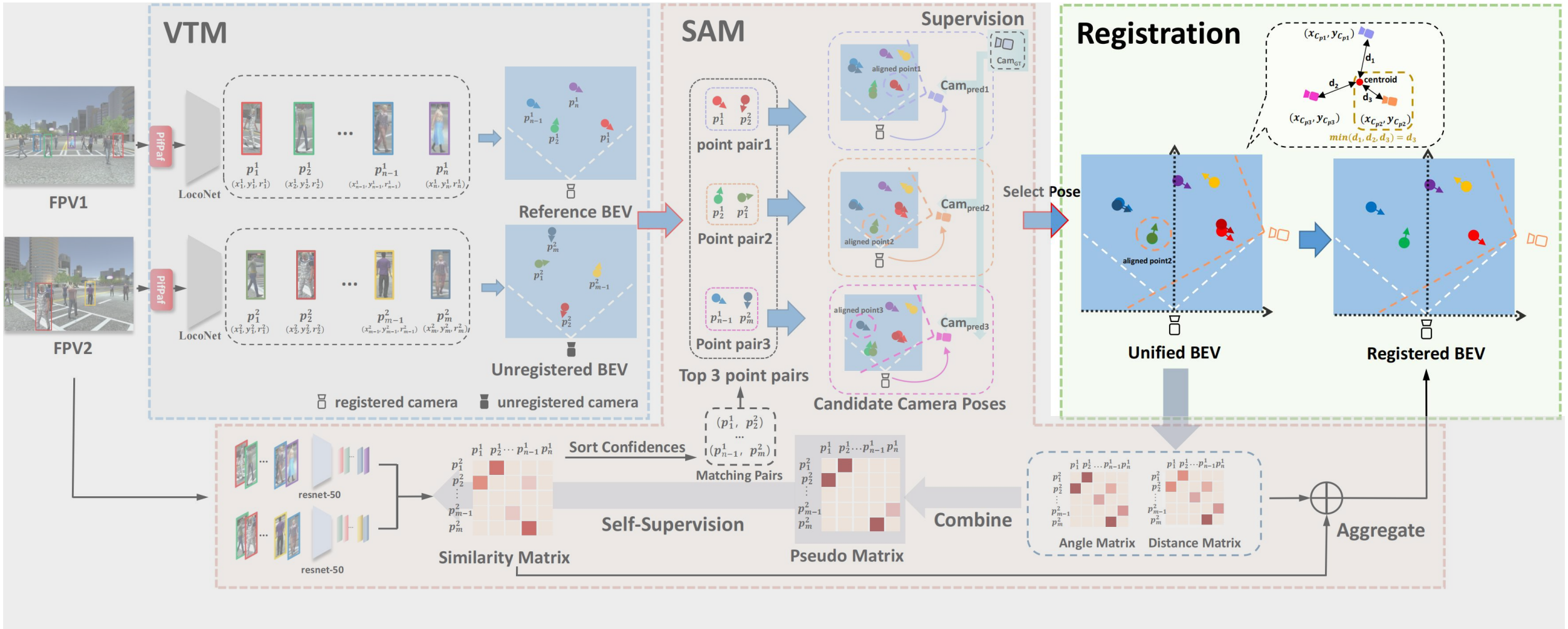
Methods


VTM
FPV1
FPV2
PifPaf
LocoNet
Reference BEV
Unregistered BEV
registered camera
unregistered camera
SAM
Supervision
point pair1
Point pair2
Point pair3
Top 3 point pairs
aligned point1
aligned point2
aligned point3
Candidate Camera Poses
Registration
Select Pose
Unified BEV
Registered BEV
centroid
resnet-50
Sort Confidences
Matching Pairs
Self-Supervision
Similarity Matrix
Pseudo Matrix
Combine
Angle Matrix
Distance Matrix
Aggregate

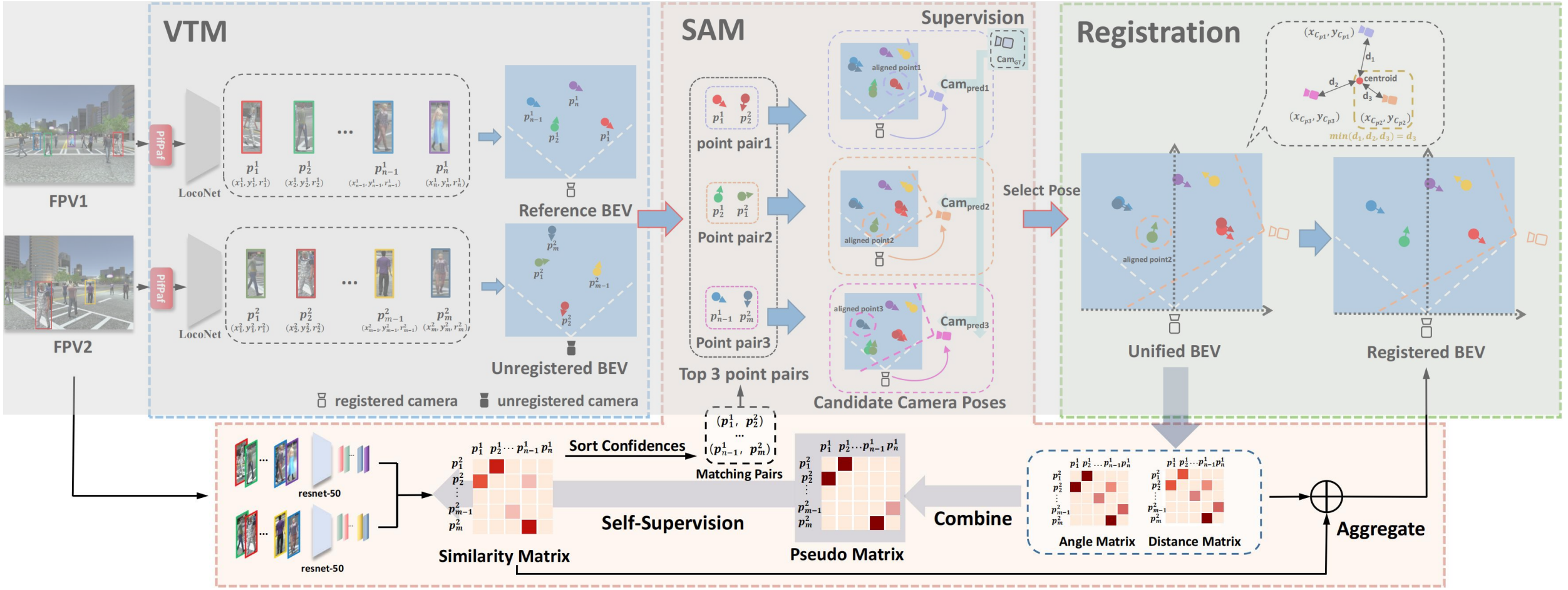
Methods


VTM
FPV1
FPV2
PifPaf
LocoNet
Reference BEV
Unregistered BEV
registered camera
unregistered camera
SAM
Supervision
point pair1
Point pair2
Point pair3
Top 3 point pairs
aligned point1
aligned point2
aligned point3
Cam_pred1
Cam_pred2
Cam_pred3
Cam_GT
Candidate Camera Poses
Select Pose
Registration
centroid
min(d1, d2, d3) = d3
Unified BEV
Registered BEV
resnet-50
resnet-50
Sort Confidences
Matching Pairs
Similarity Matrix
Self-Supervision
Pseudo Matrix
Combine
Angle Matrix
Distance Matrix
Aggregate

Methods


VTM
SAM
Supervision
Registration
FPV1
FPV2
PifPaf
LocoNet
Reference BEV
Unregistered BEV
registered camera
unregistered camera
point pair1
Point pair2
Point pair3
Top 3 point pairs
aligned point1
aligned point2
aligned point3
Candidate Camera Poses
Select Pose
centroid
Unified BEV
Registered BEV
resnet-50
Sort Confidences
Matching Pairs
Self-Supervision
Similarity Matrix
Pseudo Matrix
Combine
Angle Matrix
Distance Matrix
Aggregate

Methods


VTM
SAM
Supervision
Registration
FPV1
FPV2
PifPaf
LocoNet
Reference BEV
Unregistered BEV
registered camera
unregistered camera
point pair1
Point pair2
Point pair3
Top 3 point pairs
aligned point1
aligned point2
aligned point3
Cam_pred1
Cam_pred2
Cam_pred3
Cam_GT
Candidate Camera Poses
Select Pose
centroid
min(d1, d2, d3) = d3
Unified BEV
Registered BEV
resnet-50
resnet-50
Sort Confidences
Matching Pairs
Similarity Matrix
Self-Supervision
Pseudo Matrix
Combine
Angle Matrix
Distance Matrix
Aggregate

Experiments

CSRD-V

CSRD-R

Experiments

Table 1. Camera registration results. The top half is comparison experiments, the bottom half is ablation study, in which 'Cam.Pos.Avg' and 'Cam.Ori.Avg' present the average error in meters of the camera position and the orientation error in degrees in BEV, 'Cam.Pos@d' represents the percentage of distance error within d meters and 'Cam.Ori.@r' represents the percentage of angle error within r degrees.

Methods	Cam.Pos.Avg	Cam.Ori.Avg	Cam.Pos@0.5	Cam.Pos.@1	Cam.Pos.@1.5	Cam.Ori.@5	Cam.Ori.@10	Cam.Ori.@15
Monoloco++ [8]	3.00	21.84	7.60%	21.60%	36.40%	17.50%	34.60%	47.10%
DMHA [23]	5.99	47.43	46.50%	47.60%	48.60%	46.20%	50.00%	53.60%
SIFT [41]	7.11	144.46	1.26%	2.34%	3.60%	4.80%	8.20%	11.10%
LoFTR [54]	11.50	90.11	0.70%	1.20%	1.70%	3.70%	6.50%	8.50%
SuperGlue [48]	11.17	89.74	0.60%	1.10%	1.50%	3.70%	6.50%	8.60%
CVNet [35]	11.38	115.10	0.88%	1.25%	1.75%	3.10%	5.5%	7.40%
R2Former [65]	13.55	102.52	0.35%	0.47%	0.83%	3.90%	7.20%	9.50%
Max	2.27	15.22	20.00%	42.30%	59.60%	33.90%	60.30%	76.00%
Random	1.91	12.62	21.60%	47.30%	65.00%	37.50%	65.80%	81.20%
w/o pre-train	6.98	33.02	0.50%	1.40%	3.20%	10.20%	20.90%	29.50%
w/o GT δ_0	0.93	5.91	37.80%	71.80%	85.60%	59.10%	85.60%	94.30%
Ours	0.89	5.78	42.20%	72.40%	88.40%	59.50%	86.50%	94.80%

Table 2. Subject registration results. The expression of metrics of subject here is in the same way as Table 1.

Methods	Sub.Pos.Avg	Sub.Ori.Avg	Sub.Pos.@0.5	Sub.Pos.@1	Sub.Pos.@1.5	Sub.Ori.@5	Sub.Ori.@10	Sub.Ori.@15
Monoloco++ [8]	1.32	32.50	26.05%	61.47%	77.65%	13.21%	26.05%	38.17%
MVDetr [29]	2.41	-	11.18%	29.54%	46.07%	-	-	-
MVDet [30]	2.44	-	11.28%	29.19%	45.65%	-	-	-
w/o pre-train	6.35	89.29	1.62%	6.62%	11.41%	2.29%	4.74%	6.97%
w/o GT δ_0	0.83	16.36	41.15%	77.89%	89.31%	32.30%	56.79%	72.77%
Max	1.27	21.56	37.39%	72.38%	82.87%	30.46%	54.95%	69.13%
Random	1.06	17.19	39.19%	74.62%	85.07%	33.61%	59.01%	73.39%
Ours	0.75	14.67	43.23%	81.43%	92.12%	35.07%	63.24%	79.15%

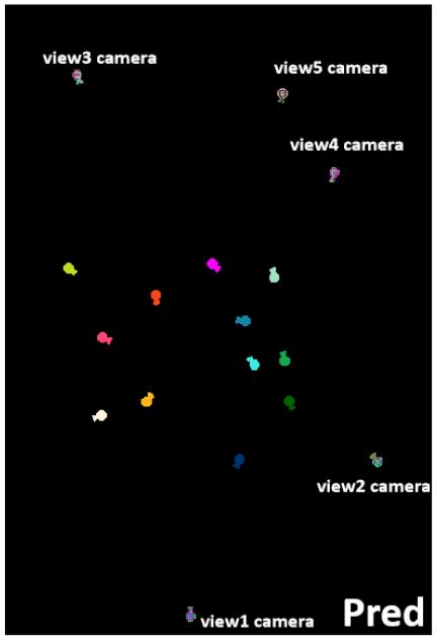
Table 3. Cross-view subject association results.

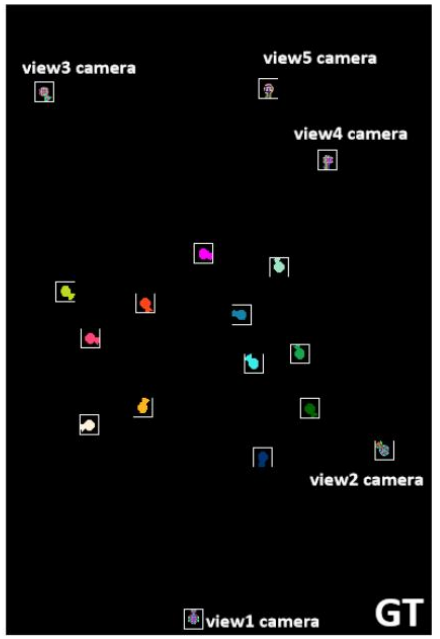
Methods	Precision	Recall	F_1
Baseline [24]	57.48%	82.98%	66.78%
Ours	79.33%	95.45%	85.98%
w GT re-id (oracle)	77.97%	98.18%	86.43%

Table 5. Results on CSRD-R for different numbers of views.

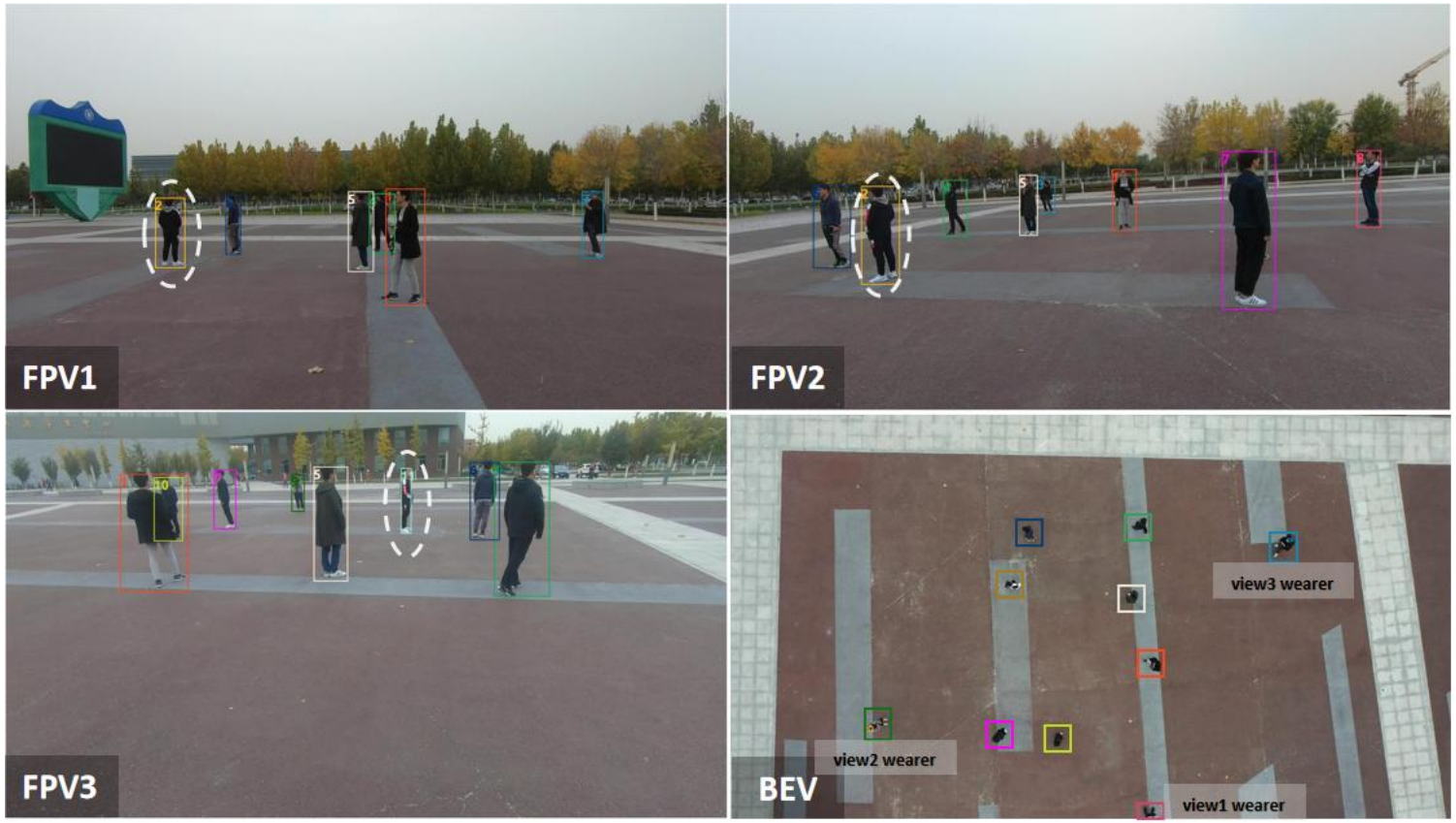
	Two Views	Three Views	Four Views
Ours	82.50%	85.07%	86.31%

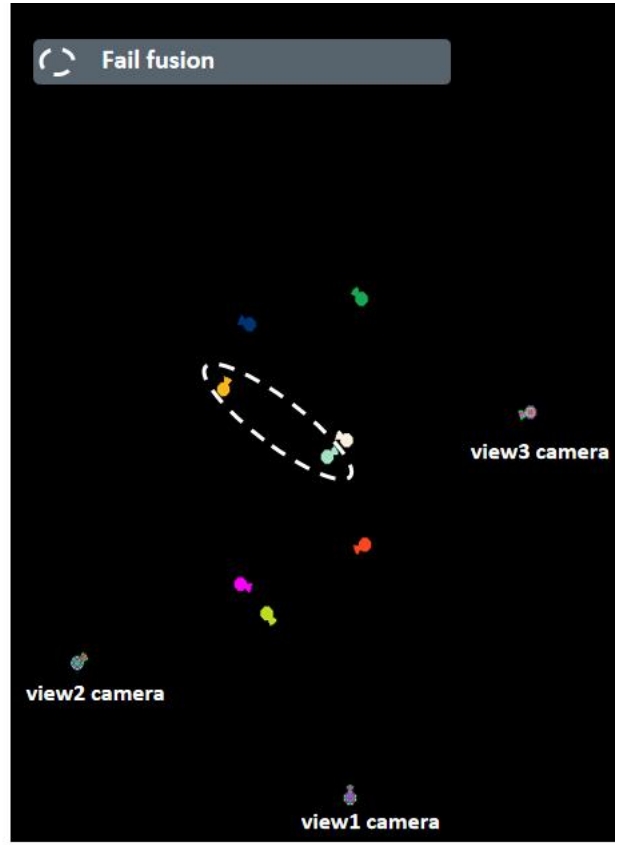
Visualization





Visualization





Thanks for your watching!