FARFusion: A Practical Roadside Radar-Camera Fusion System for Far-Range Perception

Yao Li, Yingjie Wang, Chengzhen Meng, Yifan Duan, Jianmin Ji, Yu Zhang and Yanyong Zhang, *Fellow, IEEE*

TABLE I
ABLATION STUDY ON SEQ1-3 SPLITS IN $Y = 150 - 330m$ RANGE

Seq	AS^{*}_{2D}	L_{2D}	L_{BEV}	$AS^{\dagger}_{BEV}$	$AP_{BEV}(\%)\uparrow$ $dis=1.0$	$AP_{BEV}(\%)\uparrow$ $dis=1.5$	Euc. MAE(m)↓ $dis=1.0$	Euc. MAE(m)↓ $dis=1.5$
1	✓				32.1	78.1	1.325	1.713
				✓	26.7	59.0	1.270	1.700
	✓	✓			62.1	85.0	**0.914**	1.146
	✓	✓	✓		71.0	88.7	0.919	1.080
	✓	✓	✓	✓	**71.3**	**88.9**	0.919	**1.079**
2	✓				27.7	77.0	1.345	1.770
	✓	✓			44.3	73.6	0.973	1.308
	✓	✓	✓		73.3	87.9	**0.867**	**1.017**
	✓	✓	✓	✓	**74.2**	**90.1**	0.874	1.028
3	✓				28.4	74.2	1.349	1.781
	✓	✓			29.1	53.6	**1.027**	1.409
	✓	✓	✓		62.8	85.1	1.059	**1.250**
	✓	✓	✓	✓	**62.9**	**87.4**	1.110	1.312

AS^{*}_{2D} represents the 2D association of Radar-camera points on the 2D plane. L_{2D} represents the refinement module with L_{2D} function. L_{BEV} represents the refinement module with L_{BEV} function. $AS^{\dagger}_{BEV}$ represents the BEV association of Radar-camera points on the BEV plane.

TABLE II
COMPARISONS OF DIFFERENT CALIBRATION METHODS ON SEQ1 SPLIT IN $Y = 150 - 330m$ RANGE

Method	Avg $AP_{BEV}(\%)\uparrow$ $dis=1.0$	Avg $AP_{BEV}(\%)\uparrow$ $dis=1.5$	Avg Lat. MAE(m)↓ $dis=1.0$	Avg Lat. MAE(m)↓ $dis=1.5$	Avg Long. MAE(m)↓ $dis=1.0$	Avg Long. MAE(m)↓ $dis=1.5$
DLT-four points	10.3	32.7	0.543	0.588	1.280	1.662
DLT-global points	32.1	78.1	**0.084**	**0.087**	1.319	1.708
Refinement	**71.0**	**88.7**	0.124	0.125	**0.899**	**1.062**

LiDAR provides positions for all vehicles in this range as the ground truth.

TABLE III
THE CROSS-VALIDATION EXPERIMENTS OF **M** ON SEQ1-2 SPLITS IN $Y = 150 - 330m$ RANGE

Test Splits	Refined M	$AP_{BEV}(\%)\uparrow$ $dis=1.0$	$AP_{BEV}(\%)\uparrow$ $dis=1.5$	Lat. MAE(m)↓ $dis=1.0$	Lat. MAE(m)↓ $dis=1.5$	Long. MAE(m)↓ $dis=1.0$	Long. MAE(m)↓ $dis=1.5$
Seq1	Seq1(part)	71.0	88.7	**0.124**	**0.125**	0.899	1.062
Seq1	Seq2(part)	**76.1**	**89.5**	0.139	0.145	**0.811**	**0.946**
Seq2	Seq2(part)	**74.2**	**90.1**	0.145	0.148	**0.842**	**0.997**
Seq2	Seq1(part)	66.8	86.6	**0.134**	**0.135**	0.943	1.129

APPENDIX

A. Ablation Study

We perform a set of ablation studies and summarize the results on seq1-3 splits in Tab. I. Experiments show the impact of different components of our FARFusion across various scenes. The detailed analyses can be found in the corresponding section of our paper.

B. Evaluation of Transformation Parameters Refinement

In this subsection, we closely look at the transformation parameters refinement. The experiments serve as supplemental studies for our paper.

Impact of reference points in different ranges. We employ the DLT [1] to calibrate the initial transformation parameters between Radar and pixel coordinate systems. To examine the impact of reference points in different ranges on the initial calibration, we have selected four precise pairs of fusion points and corresponding 2D points in the $Y = 150 - 330m$ range. Then we calculate the homography matrix using the DLT. In Tab. II, we see that the performance is not as good as the method that uses global points in the $Y = 150 - 500m$ range. This is because selecting only four points in the near-field range can cause the calculated homography matrix to settle into a local optimum due to the limited number of reference points. Thus, it's more effective to use global reference points in the $Y = 150 - 500m$ range for initial parameters calibration. The results also show that our refinement pipeline can enhance the detection performance with higher AP values.

Cross-validation experiments across different sequences. In our paper, we individually evaluate the refined **M** in every sequence. We have also added the cross-validation experiments of **M** on seq1-2 splits in Tab. III. We find that the reference points in various splits have different effects on the refined **M**, resulting in different detection performances of the fusion pipeline. For example, in the case of seq1 split, the **M** is refined separately on seq1 and seq2 splits. Our FARFusion utilizes the two **M** values to perform view transformation, achieving average AP_{BEV} values of 79.85% and 82.8%, respectively.

Impact of sensor motions. We have assessed the impact of sensor motions on our refinement pipeline in Tab. IV. We have added various rotation and translation noises to the initial **M** to imitate the sensor motions and generated new noisy reference points for refinement. We see that when we have $\Delta T_{xy} <= 0.5m$, our refinement can still refine the **M** successfully, achieving higher detection performance compared to when not using our refinement. However, when we have $\Delta T_{xy} = 1.0m$, i.e., a translation of 1.0m along both the X and Y axes on the BEV plane representing a significant sensor motion, it's difficult for the refinement pipeline to work effectively due to the large location deviations of reference points. Our refinement pipeline is robust against the sensor

Yao Li, Yingjie Wang, Chengzhen Meng, Yifan Duan, Jianmin Ji, Yu Zhang and Yanyong Zhang are with School of Computer Science and Technology, University of Science and Technology of China, Hefei, China (e-mail: {zkdly, yingjiewang, czmeng, dyf0202}@mail.ustc.edu.cn, {jianmin, yuzhang, yanyongz}@ustc.edu.cn). *(Corresponding author: Yanyong Zhang.)*

TABLE IV
THE IMPACT OF DIFFERENT INITIAL $\mathbf{M}$ NOISES ON REFINEMENT ON SEQ1 SPLIT IN $Y = 150 - 330m$ RANGE

Num	Refinement	ΔR_θ	ΔT_{xy}	$\mathrm{AP}_{\mathrm{BEV}}(\%) \uparrow$ $dis = 1.0$	$dis = 1.5$	Euc. MAE(m)$\downarrow$ $dis = 1.0$	$dis = 1.5$	Matching Ratio(%)
1-1		0.0	0.0	32.1	78.1	1.325	1.713	89.3
1-2	✓	0.0	0.0	71.0	88.7	0.919	1.080	88.2
2-1		0.1	0.0	24.5	70.2	1.323	1.821	89.4
2-2	✓	0.1	0.0	69.6	88.5	0.933	1.101	88.4
3-1		0.1	0.5	22.4	63.1	1.396	1.868	77.2
3-2	✓	0.1	0.5	58.4	79.4	0.999	1.207	76.5
4-1		0.1	1.0	3.30	42.7	1.575	2.040	61.5
4-2	✓	0.1	1.0	4.30	15.2	1.100	1.517	59.4
5-1		0.5	0.0	17.1	59.2	1.395	1.934	89.7
5-2	✓	0.5	0.0	51.9	83.6	1.084	1.384	89.2
6-1		0.5	1.0	7.90	41.5	1.544	1.982	63.1
6-2	✓	0.5	1.0	0.20	0.90	1.170	1.588	51.4

motions within a certain motion range.

The change of matching relation in refinement. To explicitly show the impact of our refinement pipeline on the matching relation of reference points, we have also defined the matching ratio as the proportion of the number of matching targets in 2D association relative to the total number of targets detected by the camera. We record the matching ratio values in Tab. IV. We see that our refinement pipeline will hardly change the matching relation except for the case of $\Delta R_\theta = 0.5^\circ, \Delta T_{xy} = 1.0m$, which involves a considerable noise level such that the reference points are rendered unreliable in this case. This shows that our refinement pipeline provides more accurate transformation parameters for the fusion pipeline by fine-tuning the initial $\mathbf{M}$.

REFERENCES

[1] Y. I. Abdel-Aziz, H. M. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Photogramm. Eng. Remote Sens.*, vol. 81, no. 2, pp. 103–107, 2015.