On the Applicability of General LiDAR Registration to V2X Data Alignment

Kyungmin Kim, Soonmin Hwang*
Department of Automotive Engineering, Hanyang University
Seoul, Republic of Korea
tnrud17@hanyang.ac.kr, soonminh@hanyang.ac.kr

***Abstract*—Vehicle-to-Everything (V2X) cooperative perception in autonomous driving is attracting increasing attention as a means to overcome the limitations of individual vehicle sensors and expand the perception range. Achieving this objective requires precise alignment of heterogeneous sensor information into a unified coordinate system. However, unlike conventional registration problems, V2X environments present unique challenges such as low overlap, sensor asymmetry, and limited communication bandwidth. In this study, we systematically evaluate representative alignment approaches under the V2X scenario using the DAIR-V2X dataset. The experimental results reveal that while each method shows certain advantages, they also exhibit significant limitations in terms of robustness, accuracy, and communication efficiency under realistic conditions. These findings highlight that although registration techniques are well established in general robotics, they are not directly transferable to V2X environments, motivating further research into robust and communication-aware solutions for reliable V2X data alignment.**

***Index Terms*—Vehicle-to-Everything (V2X), Cooperative Perception, Point Cloud Registration, Autonomous Driving**

I. Introduction

With the advancement of autonomous driving technology, V2X cooperative perception has gained significant attention as a key enabler for improving traffic safety and efficiency beyond the sensing limitations of individual vehicles. By sharing information from different viewpoints between vehicles and infrastructure, perception can be extended into occluded regions, enabling early risk detection. However, for such cooperative perception to function effectively, both platforms must integrate their information within a common coordinate system. This requires reliable data alignment, which is particularly challenging in real-world scenarios due to difficulties in hardware synchronization and limited point cloud overlap caused by position and field-of-view differences.

While most existing V2X cooperative perception studies have concentrated on downstream tasks such as object detection, tracking, and BEV representation learning, they typically assume that data alignment has already been achieved. In practice, however, alignment in V2X is far more difficult than in conventional single-agent setups. Sensor heterogeneity, non-overlapping viewpoints, and low point cloud overlap make alignment highly error-prone, yet these challenges are often bypassed or insufficiently addressed. Furthermore, prior studies rarely consider the critical impact of communication bandwidth and information-sharing structures, which fundamentally constrain how alignment can be realized in V2X systems.

To bridge this gap, this paper selects representative candidates across different fusion stages that are potentially applicable to V2X setups. The goal is to evaluate whether existing registration methods can robustly achieve data alignment in real V2X scenarios, and to analyze their failure cases and bandwidth efficiency, thereby providing practical insights and future directions for V2X data alignment research.

II. Existing V2X Data Alignment Approaches

In V2X cooperative perception, data alignment can be categorized into early, intermediate, and late fusion depending on the stage at which information is exchanged, as summarized in Figure 1. This categorization serves not only as a taxonomy but also as a framework to analyze the trade-off between communication efficiency and alignment accuracy. In the following subsections, we discuss the exchanged units and representative approaches at each stage.

A. Early Fusion

Early fusion exchanges raw point clouds to estimate the relative pose. The most basic approaches include classical registration algorithms such as ICP [1] and TEASER++ [5], which iteratively update correspondences between two point clouds and compute a rigid transformation that minimizes the sum of point-to-point or point-to-set distances. However, in V2X setups, the limited overlap between vehicle and infrastructure viewpoints makes these approaches highly sensitive to initialization and prone to failure. Recent learning-based methods still rely on raw point clouds as input but aim to improve robustness by learning correspondences and their confidence. Buffer-X [2], GeoTransformer [6], and PARENet [7] belong to this category. In particular, Buffer-X has shown relatively stable registration performance in low-overlap conditions and demonstrated zero-shot generalization to outdoor datasets. Nevertheless, due to the direct use of raw point clouds, the communication cost remains high, and its robustness in extremely low-overlap V2X scenarios has not yet been fully validated.

*Corresponding author: Soonmin Hwang (soonminh@hanyang.ac.kr).

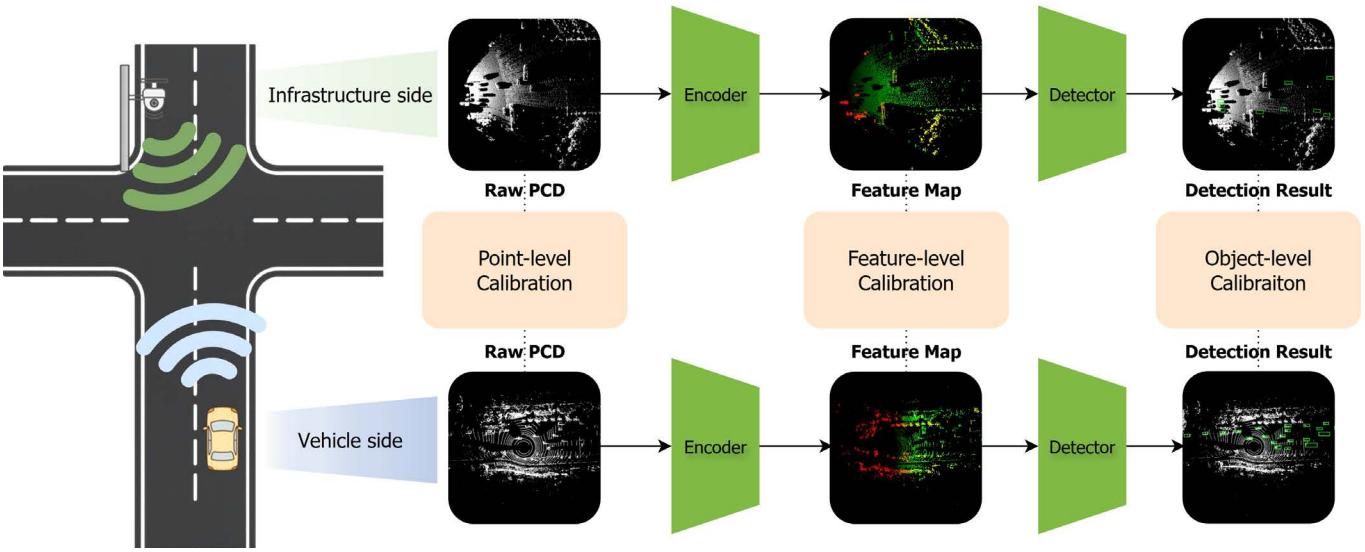

Fig. 1. Data alignment approaches in V2X cooperative perception. Early fusion (point-level) exchanges raw point clouds, Intermediate fusion (feature-level) exchanges learned feature representations, and Late fusion (object-level) exchanges object detections. Each approach exhibits distinct trade-offs between communication efficiency and registration accuracy.

B. Intermediate Fusion

Intermediate fusion exchanges feature representations instead of raw point clouds. This approach is designed to establish effective correspondences even under limited overlap. Representative methods include EYOC [8], CAST [9], and UGP [10]. However, feature descriptors that improve accuracy are often larger or more complex than raw point clouds, which conflicts with the strict communication constraints in V2X environments. For this reason, we excluded intermediate fusion from our experimental evaluation. Nonetheless, if lightweight feature representations tailored to communication constraints are developed, intermediate fusion could become a promising direction in future research.

C. Late Fusion

Late fusion exchanges object-level information for data alignment. Representative examples include V2I-Calib [11] and V2I-Calib++ [3], which match detected bounding boxes from each agent and convert box corners into feature points for relative pose estimation. Since raw point clouds are not exchanged, communication efficiency is extremely high, and this approach integrates naturally with existing detection pipelines. However, it is sensitive to detection errors such as missing objects, false positives, and localization noise, and the asymmetric viewpoints in V2X scenarios make object matching inherently unstable. Therefore, while late fusion is the most practical choice in terms of communication efficiency, its alignment accuracy fundamentally depends on the performance of object detectors.

D. Summary

Data alignment approaches can be categorized into early fusion based on raw point clouds, intermediate fusion based on feature representations, and late fusion based on object-level information. Each approach exhibits a different trade-off between communication efficiency and registration accuracy, and their applicability to V2X scenarios varies. In this work, we focus on three representative methods: the classical early fusion approach ICP [1], the learning-based early fusion method Buffer-X [2], and the object-level late fusion approach V2I-Calib++ [3]. Intermediate fusion methods were excluded due to their misalignment with communication efficiency requirements. The following sections evaluate these three approaches under V2X conditions and empirically analyze their strengths and limitations in real scenarios.

III. Experiment

A. Metrics

We adopt three standard metrics for point cloud registration:

- *Relative Rotation Error (RRE)*

$$\mathrm{RRE} = \arccos\left(\frac{\mathrm{Tr}(R_{\mathrm{gt}}^{\top} R_{\mathrm{est}}) - 1}{2}\right),$$

  where R_{gt} and R_{est} denote the ground-truth and estimated rotation matrices.
- *Relative Translation Error (RTE)*

$$\mathrm{RTE} = \|t_{\mathrm{gt}} - t_{\mathrm{est}}\|_2,$$

  where t_{gt} and t_{est} represent the ground-truth and estimated translations.
- *Success Rate (SR)*

We report two variants with thresholds $\tau_t = 2.0\,\text{m}$ and $\tau_r = 5.0°$:

$$\text{SR}_t = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[\text{RTE}_i < \tau_t]\,,$$

$$\text{SR}_{t,r} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[\text{RTE}_i < \tau_t \ \wedge\ \text{RRE}_i < \tau_r]\,,$$

where N is the total number of evaluated pairs and $\mathbf{1}[\cdot]$ is the indicator function.

B. Dataset

Experiments are conducted on the DAIR-V2X dataset [4], a large-scale vehicle–infrastructure dataset collected in real traffic environments. It provides synchronized LiDAR and camera data from both agents. Table I summarizes the LiDAR specifications.

TABLE I
SPECIFICATIONS OF LIDAR SENSORS IN DAIR-V2X

Sensor	Specification
Infrastructure LiDAR	300 beams, 10Hz, 100° horizontal FOV, −30°–10° vertical FOV, ≤ 280m range, ±3cm accuracy
Vehicle LiDAR	40 beams, 10Hz, 360° horizontal FOV, −30°–10° vertical FOV, ≤ 200m range, ±0.33° vertical resolution

C. Evaluated Methods

We evaluate representative methods at both the *point-level* and the *object-level*.

At the point level, ICP and Buffer-X are compared. ICP iteratively minimizes point-to-point distances and is evaluated with perturbations of ±5 m translation and ±5° rotation. Buffer-X learns correspondences and confidence scores directly from raw point clouds. To analyze the trade-off between accuracy and bandwidth, we reduce the transmitted points to 100%, 90%, 80%, and 70%.

At the object level, we assess V2I-Calib++, which exchanges 2D detection bounding boxes obtained from synchronized images. For data alignment, each bounding box is converted into a set of proxy points by projecting its geometric boundaries into the LiDAR space. In practice, the box corners and boundary-aligned points serve as anchors for extrinsic estimation. Since its performance depends heavily on detection quality, we additionally simulate noisy and missing bounding boxes to test robustness.

D. Implementation Details

All methods are evaluated using the official implementations and pretrained weights released by the original authors. DAIR-V2X point clouds are only reformatted into the required input types (e.g., `.bin`) without additional preprocessing. ICP experiments are performed on an Intel Xeon Gold 6548Y+ CPU, while learning-based methods are evaluated on an NVIDIA RTX A6000 GPU server.

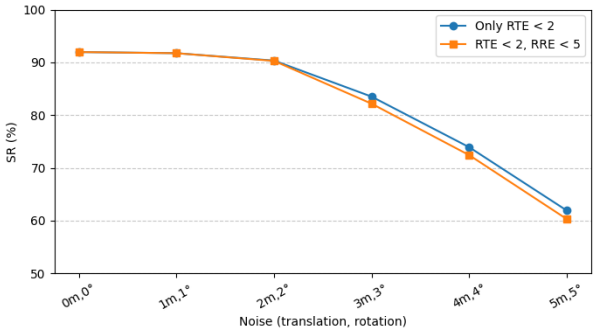

Fig. 2. SR degradation with increasing initialization noise (translation, rotation) for ICP.

IV. RESULTS

A. Quantitative Evaluation

Table II compares the three approaches (ICP, Buffer-X, and V2I-Calib++) under different initialization, noise, and transmission conditions. Metrics include RRE, RTE, SR_t, $\text{SR}_{t,r}$, as well as runtime and communication bandwidth, enabling a joint analysis of robustness and efficiency. We further provide per-method discussions.

1) ICP-based Point-level Data Alignment

Table II (top block) reports the results for ICP. Even without injected noise, SR_t only reached 91.95%, not 100%. This is mainly because some DAIR-V2X ground-truth poses contain errors of several tens of centimeters to over 1 m, making it impossible for ICP to satisfy the <2 m condition even when initialized from GT. Moreover, the large disparity in LiDAR configurations (40-beam 360° vs. 300-beam 100°) often leads to very limited overlap, amplifying sensitivity to initialization. As initialization noise increased, the success rate declined rapidly (0 m, 0° → 5 m, 5°: SR_t 91.95% → 61.91%), clearly demonstrating ICP's fragility.

2) Buffer-X Learning-based Data Alignment

Table II (middle block) shows the results for Buffer-X. Although Buffer-X is recognized for strong zero-shot performance on standard benchmarks, in V2X its SR_t dropped to 63.72% even with full point transmission. This suggests that robustness does not directly transfer under V2X-specific challenges such as low overlap and cross-sensor domain gaps.

Compared to ICP, Buffer-X is less dependent on initialization, maintaining moderate success rates even without coarse pose priors. However, absolute SR remains within the 50–60% range, limiting its reliability for cooperative perception.

As shown in Figure 3, reducing the transmitted point ratio further decreased SR (100%→70%: 63.72%→55.38%). This illustrates the accuracy–bandwidth trade-off, but the low absolute SR undermines its practical utility.

TABLE II
COMPARISON OF METHODS UNDER V2X CALIBRATION SETTINGS. SR_t: RTE< 2 M. $SR_{t,r}$: RTE< 2 M & RRE< 5°.

Init	Method	Setup	RRE (°) ↓	RTE (m) ↓	SR_t (%) ↑	$SR_{t,r}$ (%) ↑	Time (s)	Bandwidth (Mbps @10 Hz)
✓	ICP [1]	Noise: 0 m, 0°	1.1788	1.0437	**91.95**	**91.95**	5.2463	128.00
	ICP	Noise: 1 m, 1°	1.1795	1.0438	91.73	91.73	5.2731	128.00
	ICP	Noise: 2 m, 2°	1.1706	1.0459	90.33	90.27	5.4747	128.00
	ICP	Noise: 3 m, 3°	1.2720	1.0466	83.50	82.16	5.6728	128.00
	ICP	Noise: 4 m, 4°	1.3091	1.0192	73.91	72.41	5.8565	128.00
	ICP	Noise: 5 m, 5°	1.3679	1.0340	61.91	60.29	6.0176	128.00
×	Buffer-X [2]	PCD 100%	1.1911	1.1986	**63.72**	**63.49**	3.0660	128.00
	Buffer-X	PCD 90%	1.1796	1.2037	61.43	61.05	2.8853	115.20
	Buffer-X	PCD 80%	1.2059	1.2125	58.72	58.20	3.0456	102.40
	Buffer-X	PCD 70%	1.2199	1.2255	55.38	54.46	2.4882	89.60
×	V2I-Calib++ [3]	All GT boxes	1.2177	0.9831	**62.34**	**62.28**	0.3200	12.88
	V2I-Calib++	Box drop 10%	1.2650	1.0024	57.16	56.99	0.8082	11.59
	V2I-Calib++	Box drop 20%	1.3041	1.0510	51.64	51.36	0.5017	10.31
	V2I-Calib++	Box drop 30%	1.3408	1.0767	42.45	42.28	0.3001	9.02
	V2I-Calib++	Noise: 0.1 m, 1°	1.5660	1.1029	47.19	46.57	1.1639	12.88
	V2I-Calib++	Noise: 0.2 m, 2°	2.1255	1.2455	30.64	29.36	0.7805	12.88

Note. ICP was evaluated on CPU only.

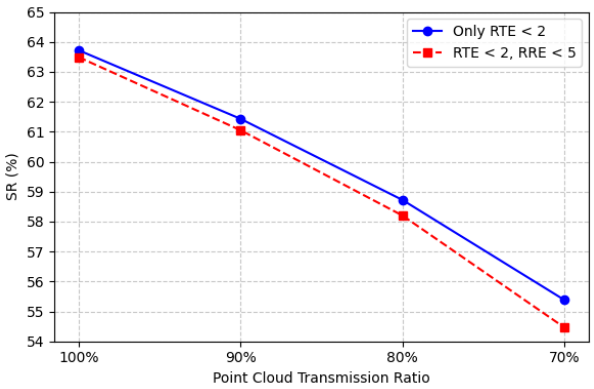

Fig. 3. Performance of Buffer-X under varying transmitted point cloud ratios. SR remains at an absolutely low level and decreases progressively as the transmission ratio is reduced.

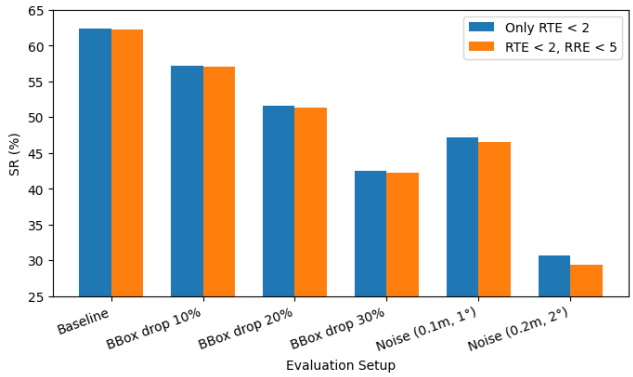

Fig. 4. V2I-Calib++ sensitivity to box drop and added noise. SR declines sharply under missing or noisy detections.

3) *V2I-Calib++ Object-level Data Alignment*

Table II (bottom block) reports results for V2I-Calib++. In the baseline setting (all GT boxes), SR_t=62.34%, comparable to Buffer-X, while requiring only 12.88 Mbps—over 10× more efficient. Moreover, when successful, V2I-Calib++ achieved higher-quality alignment (RTE=0.98 m, RRE=1.22°).

However, robustness issues are evident. As shown in Figure 4, SR degraded significantly under box drop and noise. With 30% missing boxes, SR fell to 42.45%, and with 0.2 m translation and 2° rotation noise, SR collapsed to ∼30%. This highlights strong sensitivity to both detection recall and localization noise, clarifying why late fusion has become less favored in recent literature.

Thus, V2I-Calib++ offers excellent bandwidth efficiency and precise alignment when successful, but its reliance on detection quality is a critical limitation. Enhancing robustness against missing and noisy detections is more crucial than merely improving bounding box localization accuracy.

4) *Additional Analysis on RRE and RTE*

Across successful cases, RRE and RTE remained stable. ICP and Buffer-X both maintained RTE ≈ 1.0–1.2 m and RRE ≈ 1.2°, while V2I-Calib++ achieved the lowest RTE (0.98 m) in its baseline configuration. This indicates higher alignment precision upon success, but performance degraded drastically once detection errors were introduced.

B. *Qualitative Evaluation*

Figure 5 provides a side-by-side qualitative comparison across three representative scenes, each row corresponding to a different scenario. Columns show the GT reference, ICP, Buffer-X, and V2I-Calib++ outputs, respectively. These examples illustrate typical success and failure cases that complement

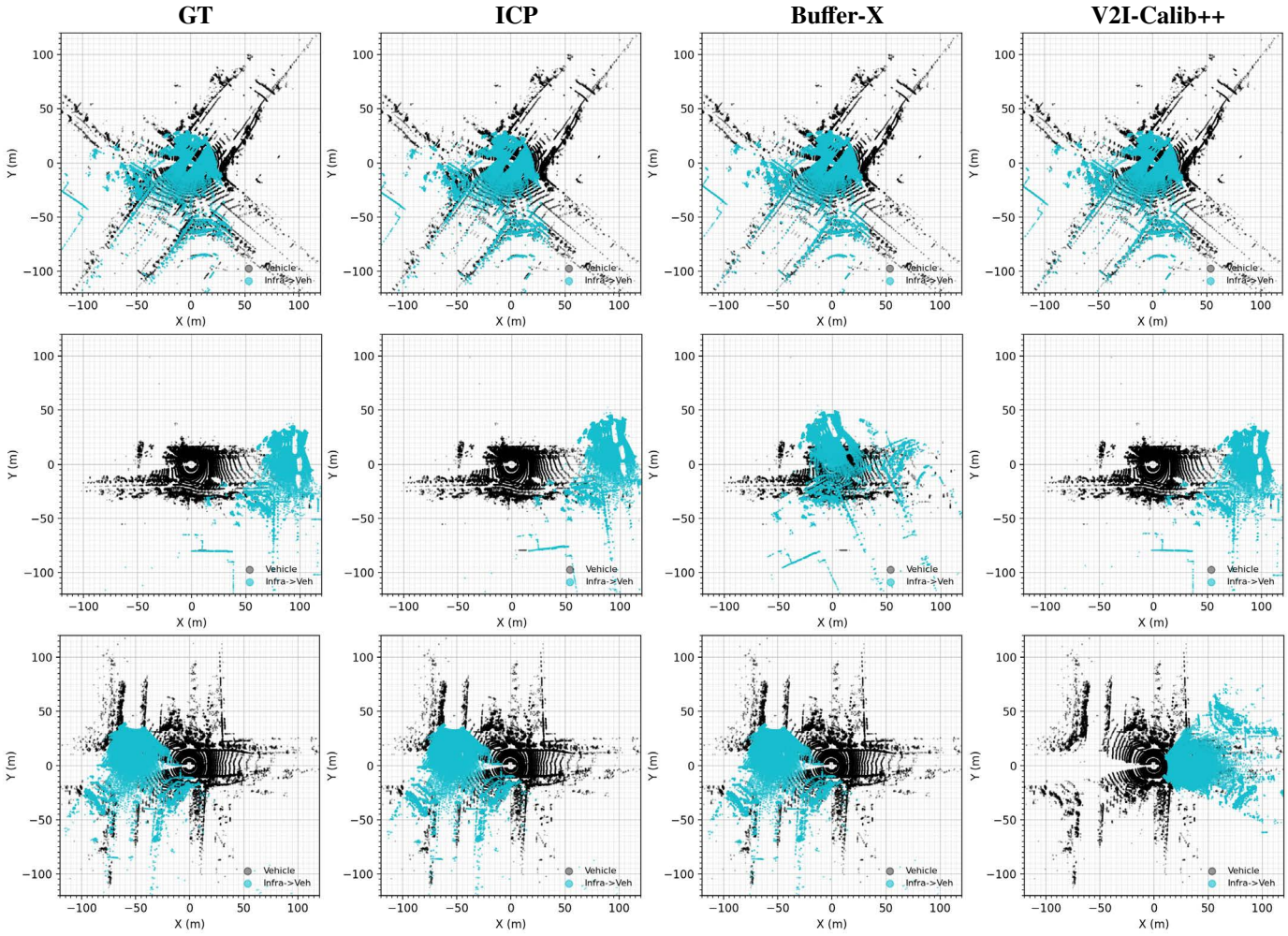

Fig. 5. Qualitative comparison across three representative scenes (rows) and four methods (columns). Columns: (A) GT, (B) ICP, (C) Buffer-X, (D) V2I-Calib++. Rows: (a) all good with slight GT jitter, (b) low overlap where V2I-Calib++ succeeds while ICP/Buffer-X fail, (c) high overlap where V2I-Calib++ fails due to the scarcity of common dynamic objects, resulting in insufficient correspondences despite abundant geometric overlap. Major/minor grids help visual assessment of residual misalignment.

the quantitative analysis.

1) ICP-based Point-level Data Alignment

ICP is highly sensitive to both ground-truth jitter and geometric overlap. In Row (a), where the GT itself contains small pose errors, ICP sometimes produces alignments that appear even better than GT, reflecting that DAIR-V2X "GT" is a refined estimate rather than an absolute reference. In Row (b), ICP collapses into local minima due to insufficient overlap, consistent with the sharp SR degradation observed in Table II. When overlap is sufficient (Row (c)), ICP achieves almost perfect registration, visually indistinguishable from GT. These observations visually support the sensitivity and SR degradation trends reported in the quantitative section.

2) Buffer-X Learning-based Data Alignment

Buffer-X maintains moderate robustness even without explicit initialization. As seen in Figure 5, when overlap is adequate, its alignment quality is comparable to GT and ICP (Rows (a) and (c)). However, in ambiguous or low-overlap scenarios (Row (b)), Buffer-X often diverges catastrophically, producing large offsets rather than gradual drift. Thus, while less dependent on initialization than ICP, Buffer-X tends to fail in a more catastrophic manner under challenging V2X conditions, which raises concerns for safety-critical use.

3) V2I-Calib++ Object-level Data Alignment

V2I-Calib++ leverages object correspondences, allowing it to succeed in cases where point-based methods fail. For example, as illustrated in Figure 5 Row (b), it successfully aligns scenes with very limited geometric overlap, where ICP and Buffer-X collapse. However, as shown in Row (c), it fails even under abundant geometric overlap, since only a few dynamic objects are jointly observed, leaving the solver with insufficient correspondences. Figure 6 provides a close-up of this Row (c) failure, including the detected bounding boxes: the infrastructure-side LiDAR (left) and vehicle-side LiDAR (right) share very few commonly visible dynamic objects,

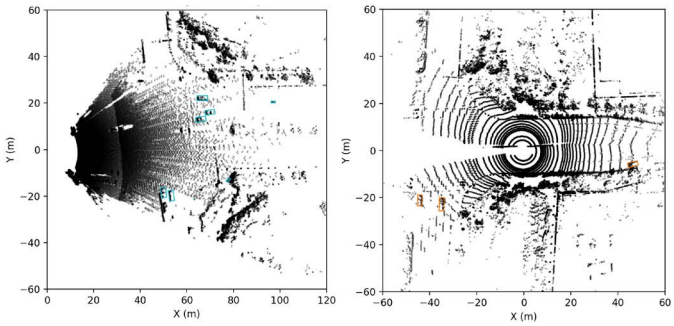

Fig. 6. Close-up analysis of the Row (c) failure in Figure 5, including the detected bounding boxes. Left: infrastructure-side LiDAR; Right: vehicle-side LiDAR. Despite large geometric overlap, the two views share only a few commonly observed dynamic objects, resulting in insufficient correspondences and causing V2I-Calib++ to misalign.

making object-level matching underconstrained. By contrast, point-level methods can still exploit rich static geometry (e.g., façades, road edges), explaining why they succeed in the same case. Together, these observations underscore the fragility of V2I-Calib++ to detection sparsity.

4) Discussion

Taken together, the qualitative analysis highlights complementary properties: (i) ICP achieves excellent accuracy when overlap is sufficient but collapses otherwise. (ii) Buffer-X generalizes better without initialization but fails catastrophically under difficult conditions. (iii) V2I-Calib++ can succeed even under low overlap but is brittle in sparse-object or noisy-detection scenarios. These case studies visually explain the quantitative trends and suggest that practical V2X data alignment requires balancing geometric overlap, initialization priors, and detector robustness.

V. Conclusion

This work presented a comprehensive evaluation of representative calibration approaches for V2X cooperative perception using the DAIR-V2X dataset. By analyzing early fusion (ICP), learning-based point-level registration (Buffer-X), and late fusion (V2I-Calib++), we identified the unique strengths and limitations that arise in realistic V2X settings.

The experiments show that traditional registration techniques, though effective in conventional robotics, are not directly transferable to V2X environments due to challenges such as low overlap, sensor asymmetry, and dependency on detection quality. ICP achieves accurate alignment only under favorable initialization and overlap, Buffer-X alleviates initialization dependence but remains limited in robustness, and V2I-Calib++ offers superior bandwidth efficiency but is fragile against noisy or missing detections. These findings highlight that no existing approach alone can fully address the requirements of reliable cooperative perception.

Looking ahead, future research should focus on (i) reducing detection dropouts, (ii) ensuring robustness to noise and imperfect ground-truth references, and (iii) bridging domain gaps between heterogeneous sensors. In particular, feature-level alignment (intermediate fusion), which has recently gained attention, offers a promising direction to mitigate the weaknesses of point-level and object-level methods. Advancing this paradigm will be essential to achieving robust and scalable V2X cooperative perception.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00409492).

References

[1] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611, pp. 586–606, SPIE, 1992.

[2] M. Seo, H. Lim, K. Lee, L. Carlone, and J. Park, "BUFFER-X: Towards Zero-Shot Point Cloud Registration in Diverse Scenes," *arXiv preprint arXiv:2503.07940*, 2025.

[3] Q. Qu, X. Zhang, Y. Cheng, Y. Xiong, C. Xia, Q. Peng, Z. Song, K. Liu, X. Wu, and J. Li, "V2I-Calib++: A Multi-terminal Spatial Calibration Approach in Urban Intersections for Collaborative Perception," *arXiv preprint arXiv:2410.11008*, 2025.

[4] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, *et al.*, "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21361–21370, 2022.

[5] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 314–333, 2020.

[6] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, S. Ilic, D. Hu, and K. Xu, "GeoTransformer: Fast and Robust Point Cloud Registration with Geometric Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9806–9821, 2023.

[7] R. Yao, S. Du, W. Cui, C. Tang, and C. Yang, "PARE-Net: Position-Aware Rotation-Equivariant Networks for Robust Point Cloud Registration," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 287–303, 2024.

[8] Q. Liu, H. Zhu, Z. Wang, Y. Zhou, S. Chang, and M. Guo, "Extend Your Own Correspondences: Unsupervised Distant Point Cloud Registration by Progressive Distance Extension," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20816–20826, 2024.

[9] D. Lee, J. Lee, and J. Choi, "CAST: Cross-Attention in Space and Time for Video Action Recognition," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, pp. 79399–79425, 2023.

[10] Z. Zeng, Q. Wu, X. Zhang, L. Y. Wu, P. An, J. Yang, J. Wang, and P. Wang, "Unlocking Generalization Power in LiDAR Point Cloud Registration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22244–22253, 2025.

[11] Q. Qu, Y. Xiong, G. Zhang, X. Wu, X. Gao, X. Gao, H. Li, S. Guo, and G. Zhang, "V2I-Calib: A Novel Calibration Approach for Collaborative Vehicle and Infrastructure LiDAR Systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 892–897, 2024.