Privacy-Aware Sharing of Raw Spatial Sensor Data for Cooperative Perception

Bangya Liu
University of Wisconsin-Madison
Madison, Wisconsin, USA

Chengpo Yan
University of Wisconsin-Madison
Madison, Wisconsin, USA

Chenghao Jiang
University of Wisconsin-Madison
Madison, Wisconsin, USA

Suman Banerjee
University of Wisconsin-Madison
Madison, Wisconsin, USA

Akarsh Prabhakara
University of Wisconsin-Madison
Madison, Wisconsin, USA

Abstract

Cooperative perception between vehicles is poised to offer robust and reliable scene understanding. Recently, we are witnessing experimental systems research building testbeds that share *raw spatial sensor data* for cooperative perception. While there has been a marked improvement in accuracies and is the natural way forward, we take a moment to consider the problems with such an approach for eventual adoption by automakers. In this paper, we first argue that new forms of privacy concerns arise and discourage stakeholders to share raw sensor data. Next, we present SHARP, a research framework to minimize privacy leakage and drive stakeholders towards the ambitious goal of raw data based cooperative perception. Finally, we discuss open questions for networked systems, mobile computing, perception researchers, industry and government in realizing our proposed framework.

CCS Concepts

• **Networks** → *Network privacy and anonymity*; • **Computing methodologies** → *Vision for robotics*; • **Computer systems organization** → **Sensor networks**.

Keywords

Cooperative Perception, V2V Network, Privacy, Novel View Synthesis, Vision Foundation Model, Simultaneous localization and mapping

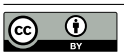
ACM Reference Format:
Bangya Liu, Chengpo Yan, Chenghao Jiang, Suman Banerjee, and Akarsh Prabhakara. 2026. Privacy-Aware Sharing of Raw Spatial Sensor Data for Cooperative Perception. In *The 27th International Workshop on Mobile Computing Systems and Applications (HotMobile '26), February 25–26, 2026, Atlanta, GA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3789514.3792039

1 Introduction

Cooperative perception is an emerging networked-systems application domain that involves sharing spatial sensor data between vehicles to enhance vehicle capabilities, safety, and user experience. As self-driving moves to higher autonomy levels, reliability is key. Cooperative perception, done correctly, can enable reliable machine vision detecting events ahead-of-time by extending the range of perception in both normal and adversarial weather. For example: An accident detected on a route can be used to reroute other vehicles. Black ice detected at a corner can be used to warn all following vehicles to turn on traction control. Sharing of such locally processed data ("events", bounding boxes, semantic labels) has been the focus of the initial wave of research and commercialization. However, *raw spatial data* (camera, lidar and radar) sharing between vehicles is touted to bring the next wave of benefits from cooperative perception where the "whole is greater than the sum of its parts" is realized. For example: lidar return from small surface area objects like pedestrians yield only a handful of points, but when stitched with other vehicle's lidars, the densified point cloud can robustly detect critical objects. Using raw data as input to learning models has dramatically improved perception [21, 24, 39, 41, 44]. We refer to data as "raw data" even though the actual communicated message is a compressed/encoded version, as long as it is not represented in a lossy style (like "events" or bounding box).

Realizing the power of raw spatial data demands innovation in perception, systems to realize high throughput, low latency, and methods to guarantee data confidentiality, integrity and availability. The expectations on all fronts are pushed to extreme levels than for processed data sharing. Recent works including RAO [44] and RECAP [32] have extensively discussed the challenges within data synchronization and point cloud registration. While there are open challenges that are being addressed in computer vision, networking and mobile computing, we jump hoops to consider the eventual practical barrier: **privacy of raw data streams**. This problem is rarely considered by recent research, but we argue that it is of utmost importance for adoption by networked vehicles. We begin by systematically showing why this is a first-order concern.

Today, auto manufacturers have privacy agreements with consumers that decide how data generated in their vehicle can be used. A single auto manufacturer can even share processed data within its ecosystem to generate timely safety alerts (e.g: Mercedes E & S class). However, because of the diversity and the number of players in the auto industry, the vision of cooperative perception can be fully realized only when *multi-automaker collaboration* is possible. An obvious challenge that emerges with multi-automakers is the mismatch in privacy agreements. One auto's consumer may opt-in for sharing data to third parties for added services, another may opt-out. Makers could also have different fine print in their agreements. So, how can the shared data be used? Should it be held to



HotMobile '26, Atlanta, GA, USA

ACM ISBN 979-8-4007-2471-8/2026/02
https://doi.org/10.1145/3789514.3792039

ego-vehicle (local) or collaborator's standards? Beyond this, new forms of privacy concerns unique to raw spatial data emerge as follows:

Participating vehicle's location: Typically, when spatial data is shared, for stitching purposes, accurate pose associated with the data frame composed of global location (from GPS) and a finer pose estimation (from local processing) is shared [18, 42, 43]. This pose data (used even with processed data sharing) would be shared under pseudonyms in line with privacy-preserving feature of vehicular communication standards [16]. At first glance, this seems to reduce the utility of data to be short-lived and not linkable over time. However, just raw spatial data alone is rich enough to extract pose (through visual localization) [28, 36] and identify vehicles (through sensor noise, driving patterns, routes) over long-terms [8, 11]. Today's vision foundation models like VGGT [37] and MapAnything [19] make it easy to extract such data. The receiving vehicle could monetize these patterns with third-party insurance or advertising companies.

Intellectual property of sensor design: A huge concern for auto manufacturers is that when raw data is shared, the surface area for reverse engineering to learn trade secrets about low-level sensor quality is greatly exposed. Moreover, as mentioned above, despite pseudonyming, these low-level properties can be linked to automaker, auto type, sensor vendor etc. over long terms. Several billions of dollars are spent on research and development of spatial sensor hardware and integration. Beyond cost, sensors also vary greatly in quality. For example, one automaker may choose a camera vendor that offers superior stabilization as part of low level IP, another may have inferior stabilization. This begs the question: why should anyone participate in raw data sharing to give away secrets on the quality of their data?

Concerns of such a high magnitude have made several automakers, auto consumers, and the public quite skeptical about raw data-based cooperative perception [4, 6, 7, 22]. This paper presents SHARP[1], a framework to alleviate privacy concerns and skepticism, and to encourage investment in all aspects (algorithms and systems infrastructure) relevant to cooperative perception with high fidelity, raw data.

Rather than dealing with these concerns at the cross-automaker privacy agreement level, the first question to ask is if past privacy-preserving computation frameworks (Secure Multi-Party computation and Fully Homomorphic encryption [2]) are suitable to limit data sharing to specific functions. Although it is theoretically feasible to represent raw spatial data operations as additions and multiplications, the high number of samples in cameras, lidars and radars, and the communication overhead for decentralized operations render such options to be slow for real-time cooperative perception tasks with 10s-100s of milliseconds tolerances.

SHARP's agenda is twofold. First, we propose a location obfuscation design that makes it challenging for raw data recipients to extract true auto location. Our design is inspired by the dramatic recent shift in 3D vision foundation models. Second, we look for inspiration in other industries where competitors have shared raw data for mutual benefit towards a shared goal. While safety can be a unifying cause (example: telcos should route all E911 calls, not

[1]SHARP: SHAring Raw spatial sensor data Privately

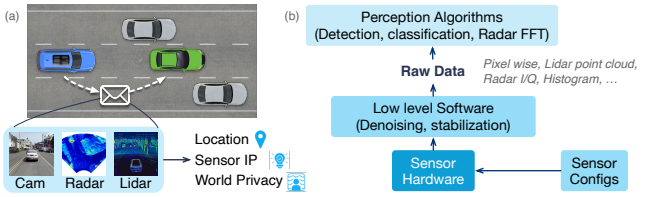


Figure 1: (a) shows the privacy leakage due to raw data streaming from blue car to the green car; (b) abstracts typical sensor architectures and shows raw data types

just their customers), there are ways (like processed data sharing) to achieve simple metrics of safety without IP data leakage. To reap the benefits of raw data, we argue and lay out the need for a coordinated effort in defining open standards that preserve IP, multi-vendor agreements, and cost models.

SHARP demonstrates our line of reasoning, backed by a feasibility study, towards solving this major problem. Our hope is to enhance awareness of this seldom cared-for, last mile problem to multiple stakeholders and thereby spurring designs of privacy-cognizant holistic solutions. We note that raw data sharing needn't be the de-facto standard for cooperative perception (with varying network capabilities), but should the infrastructure allow it and the environment demands it, it is imperative that we tackle this.

2 The Need for Raw Data

Raw spatial sensors surpass other low bandwidth data types in cooperative perception. Let us look at advantages that only raw data sharing can offer.

Cameras: Visible light cameras are sensitive to lighting and weather, and natively lack depth understanding. Cooperative perception can compensate for these by sharing (1) specific objects-of-interest (e.g: accidents) that are already detected locally by algorithms running on pixel-wise data; (2) pixel-wise image data to yield enhanced 3D understanding. Structure-from-motion, photogrammetry, neural 3D understanding all leverage pixel-wise data [13, 26, 27, 30, 33], which offers higher fidelity perception and is widely used.

Lidars: Lidars are sparser than cameras. Conventionally, the rawest level of data made available by vendors is point clouds obtained from the first return from specific laser directions. So, the lightest way to share lidar data is bounding boxes on 3D point cloud, ETSI standards [12] for example. However, for weak objects or cars in foggy conditions where the range is reduced, detection is hard with only a few spatial points. With point cloud sharing or full time delay histograms available with SPAD lidars, we can obtain dense aggregated point clouds and thus, robust object detections.

Radars: Solid-state radars are orders of magnitude sparser than cameras and lidars. Typical processing involves thresholding of raw histogram data to convert to sparse point clouds. The first level is the accumulation of point clouds that leads to the densification of objects and better object detection [3]. Second, we have exchange of intensity-only raw data. Third, phase-sync aggregation (aka coherent aggregation) can dramatically boost the detectability of weak objects such as a pedestrian standing next to a bright object like a stop sign. Here, we consider I/Q data and unthresholded, intensity-only data as "raw radar data".

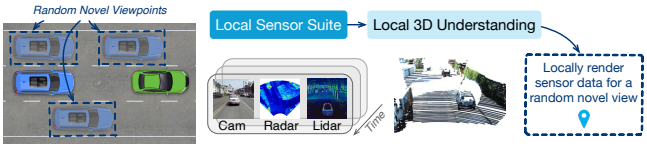


Figure 2: SHARP's approach to hide true location: use *fast* 3D understanding and rendering to constantly spoof realistic, raw data from a random view point.

We note that low-level signal processing blocks (see Fig 1) might be attached to sensors and only then expose raw data. What we refer to as "raw data" also varies from sensor-to-sensor. It is clear that there is a data hierarchy. Raw data is at the top of this hierarchy, offering the highest fidelity and enabling algorithms that enhance the degree of reliability.

3 SHARP's Design

In this section, we alleviate privacy concerns in Sec. 1.

3.1 Tackling vehicle's location leakage

Sharing raw spatial streams (cameras, lidars, radars) inevitably enables recipients to reconstruct sensor and vehicle poses. Our key insight: generating sensor measurements from alternate viewpoints protects the true location (Fig 2). By constantly randomizing viewpoints, true location tracks remain hidden even when pseudonyms are constant.

Preliminary simulation: To quantify the feasibility of this privacy paradigm, we conduct a large-scale simulation across all 73 scenes in the OPV2V [40] dataset, with 50 rollouts per scene. For each rollout, we randomly select one vehicle as the ego vehicle (i.e., the receiver) while surrounding vehicles (i.e., sharers) share perception data at forged locations—coordinates perturbed by Gaussian noise. We then evaluate the privacy level of these sharers from two perspectives: (1) the deviation of forged trajectories from ground truth trajectories, and (2) the likelihood that the ego vehicle can correlate a sharer's identity with a physical vehicle observed in the environment. Within the ego vehicle, we deploy a nearest neighbor algorithm for tracking sharers and employ the Hungarian algorithm for matching between inferred and physical trajectories. We use RMSE and N-to-N matching error rate (confusion rate) as evaluation metrics, with results presented in Fig.3. The results reveal a clear trend: as the offset between forged and ground truth positions increases, the ego vehicle's ability to identify sharers diminishes. Specifically, at an offset of 12 meters (equivalent to 4 lanes of line-of-sight distance), the confusion rate reaches 25% and the overall RMSE exceeds 45 meters. This large-scale simulation demonstrates SHARP's effectiveness in protecting sharer privacy.

Challenges: Although obfuscation can be achieved, such a solution is valid only if the sensor measurements generated are realistic. Novel View Synthesis (NVS) involves 3D reconstruction from local sensor measurements and rendering novel views. Approaches range from classical photogrammetry [26] and SLAM [33] to recent neural methods [13, 27]. However, 3D reconstruction takes seconds to hours—incompatible with automotive perception at 10+ Hz. Context-specific spoofing [31] lacks generalizability, necessitating full 3D understanding.

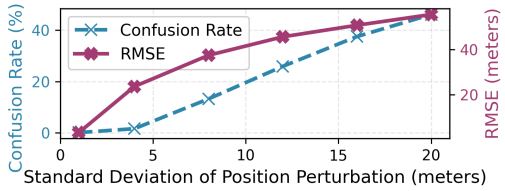


Figure 3: Preliminary simulation on the OPV2V scenes.

Our approach: Fortunately, recent breakthroughs like Dust3r [38] and Visual Geometry Grounded Transformers (VGGT) [37] have revolutionized 3D understanding, reducing reconstruction time to milliseconds—orders of magnitude faster than prior work. Rather than iterative optimization, large-scale feed-forward models perform fast inference from 2D images to 3D points, similar to monocular depth estimation [14] but with dramatically improved quality through massive training data and model scale. This paradigm shift enables low-latency automotive applications. We propose vehicles to use VGGT-like models [37] for local 3D understanding, and then render novel views at randomized locations in real time. From our simulation, such an obfuscation creates confusion in extracting the true vehicle pose. Sec 4 shows the feasibility and realism of the novel views. VGGT requires extensions for comprehensive automotive deployment. Sec 5 discusses open challenges, including occlusion-aware viewpoint selection.

While novel view based obfuscation masks exact trajectories they can still reveal general location (e.g., an intersection). To prevent receivers from building long term associations of general locations, we need to make it more challenging to map time-varying pseudonyms to uniquely identifiable sensor traits. Sec 3.2 masks low-level sensor details and prevents building long term location tracks based on general locations.

3.2 Tackling IP leakage

IP leakage from sharing raw spatial sensor data is so serious a concern that it could just render all research efforts towards raw data based cooperative perception moot.

Tiered relationship in automotive markets: In experimental testbeds built with development kits [44], one can access raw data. However, the auto industry is layered as tier 1, 2, & 3 depending on the level of interaction with the final automaker. The vast majority of final automakers only have access to processed sensor data from tiered vendors due to IP protection. Vertical integrators like Waymo and Tesla, either make sensors in-house or have special contracts to get hardware from non-competitors with access to raw data (Samsung, LG) for building their own proprietary perception. In a competitive scenario, why should one with access to raw data, share and leak secrets? Well, Sec 2 has shown significant benefits, so this is a utility-privacy trade-off.

Strawman solutions: Approaches like neural encoders are insufficient, because to use raw data, a trained decoder [39] at the receiver is needed and one would then have low-level sensor access. We could also consider cryptographic one way functions. But, they are not designed with the intention to preserve spatial correlations. So, the receiver would need to convert back to raw data before running cooperative perception. Another solution could be to add noise

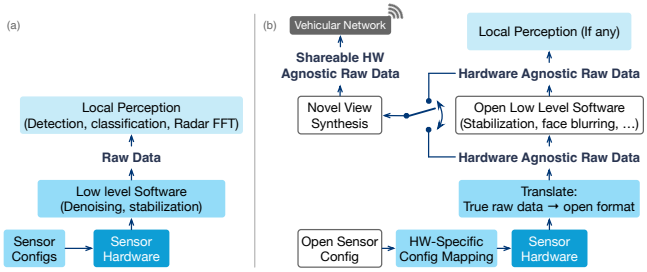


Figure 4: SHARP's approach to minimize IP leakage: open sensor configurations, open low-level software and standardizing hardware agonstic raw data as output from sensor hardware. Blue blocks are proprietary. We schedule (b) to create shareable raw data, else we schedule a proprietary stack (a).

to the raw data to make it confusing for IP theft. However, each hardware has different capabilities (sampling rate, pixel / antenna count, sensitivity etc.), and metadata about these capabilities should be shared for the recipient to fully use the raw data. This would directly give away critical data.

Our approach: We call for a common unifying representation that all automakers (with raw data) can agree on. To this end, we argue that a open low-level stack from just above the sensor hardware till "raw data" must be executed to arrive at a common representation. We envision that such a stack would be built using public knowledge. It is quite possible that such a stack will be inferior to running a proprietary stack. But atleast, such data can be shared without any IP concerns. Thus, we propose a common stack for the purpose of raw data cooperative perception.

To fully execute an open standard, one needs not only an open low-level signal processing block but also the ability to control sensor configs and hardware. For example, even though a vehicle may have 10 antennas, the open config could expose only 1 antenna's data – protecting the vehicle's secret about number of antennas. Standardizing hardware is simply infeasible! Modifying sensor configs and running a low-level software block are also inconvenient, because each vehicle may want to run their proprietary stack regardless – for local perception.

Our idea is to offset the inconvenience by carefully swapping from proprietary stack (Fig 4a) to the open stack (Fig 4b) at a duty cycle. In designing a suitable scheduling algorithm, the key constraints to consider are computing power, latency arising as a result of swaps, and the overall end-to-end latency needed for cooperative perception reaction times in challenging environments (bad weather, etc.). For example, one can switch to open configs and run open stack every once in 200 ms (5Hz) or based on demand, whereas the proprietary stack runs every 50 ms (20 Hz).

Standardization: To implement this, we still need to deal with hardware IP. Hardware varies widely in capabilities, and sensor vendors often tightly integrate hardware, configs and low level software. While configs and low-level software can rely on public, open stacks, we need an interface to various hardware. Thus, we have arrived at a bottleneck which can only be resolved via standardization. We envision bringing different hardware sensor manufacturers together to implement a *compact proprietary layer* to map their true hardware output to an agreed upon, *hardware-agnostic, open format*. On top of this, other features can also be implemented on an open stack (see Fig 4). For example: even if a radar uses a high sampling rate, a sensor vendor could modify the data to only satisfy higher-level open format requirements, on radar range and resolution. Other hardware variations to be addressed in the proprietary layer include waveform parameters, effect of sensor noise, mosaicing, size, field of view, gain, number of lidar beams, number of antennas etc. Similarly, the manufacturer would also need an interface to map the open sensor configurations to the hardware capabilities. This calls for efforts to (1) create a taxonomy and represent high-level hardware capabilities and define shareable open formats, (2) build (proprietary) software to translate open configs to native hardware, (3) build (proprietary) software to translate true raw data to open format.

Method	success rate	pose err(rmse)	time(s)/frame
VGGT	183/183	0.6900	0.089
DROID-SLAM	183/183	1.2982	0.166
COLMAP	43/183	2.8477	5.150

Table 1: A receiving car can easily estimate camera pose from received raw images with VGGT.

Cost models: Despite the need to share raw data for mutual safety, a stack swapping technology solution is likely to face some opposition (for resource-constraints reasons), unless we incentivize it as a revenue-generating stream. We argue for a cost model that a recipient pays for if they subscribe to raw data. Since the stack is open and based on publicly available information, we can treat each data to be of identical value. The total cost would be based on how frequently the recipient asks data, and any priority levels that they expect to be served at. For example: in harsh weather or when a vehicle is behind a long truck on a single lane highway, that is, when raw data need is extreme, we expect to switch priority to higher levels. We expect multi-party agreements facilitated through a lightweight, common billing system.

By scheduling stack swapping, creating standards for an open stack and incentivizing automakers with additional revenue stream, SHARP facilitates raw data exchange. As we build this, we should also work towards integrating it with default automotive embedded workflows.

4 Feasibility

Here, we show the privacy problem with raw data and feasibility of our approach. We run experiments on OPV2V [40] driven by Carla simulator [10] with Nvidia V100S 32GB.

Vehicle pose could be easily inferred from raw data: We perform a demonstrative experiment that estimates the camera poses from streams of unconstrained RGB images. 183 streams with 10 frames each. VGGT as a vision foundation model can predict camera poses and depth maps simultaneously. Table. 1 shows the fast inference, and accurate pose from VGGT compared to traditional method like COLMAP [30] as well as the SoTA visual SLAM work, DROID-SLAM [34], confirming that location privacy is a serious concern.

SHARP's novel view synthesis to obfuscate vehicle location is effective: As proposed in Sec 3.1, contributor performs 3D scene

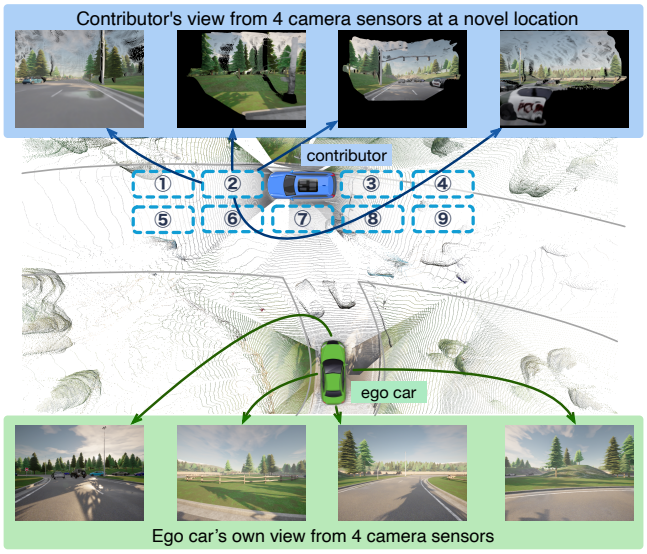


Figure 5: A demo on Town06 of OPV2V dataset, where contributor pretends it is at location 2.

Shared Data	@1	@2	true	@3	@4
MSE	0.0108	0.0083	0.0088	0.0089	0.0117
No Coop	@5	@6	@7	@8	@9
0.0142	0.0127	0.0101	0.0086	0.0117	0.0103

Table 2: Relative depth error when ego agent does no cooperative perception and when sensor data from different viewpoints are shared by contributor.

understanding with local data via vision foundation model, then renders sensor data from a novel view, and only such novel view is shared to an ego-agent. The rendering could be achieved by either plain point projection or through 3DGS [20] to reduce pixel holes. The ego agent then leverages the received views and its own sensor data to understand the surrounding. Fig 5 shows a qualitative result. Table 2 shows the overall relative depth estimation error [2]. Ego car achieves a lower error with cooperative perception, benefiting from shared data that is either true or from a novel view. Across all 9 locations, the error is comparable and generally closer locations yield lower errors.

Attacks: One concern with novel views is the potential risk of recovering how far away the novel view is from real pose based on the quality of shared rendered images (holes, missing pixels etc). It could be resolved by 3D reconstruction over a longer context. For a frame sequence from nuScenes [5] dataset, we have observed that the longer sequence we use to reconstruct the 3D point cloud, the fewer missing pixels when we perform NVS at an adjacent viewport, via a plain point cloud reprojection, as shown in Fig.6. One could mitigate this by applying random masks over the rendered images (losing some 3D information but protecting any leakage of location) or using inpainting methods [29].

[2]Output of VGGT is non-metric, hence popular benchmark on object detection is not a suitable evaluation task in this case.

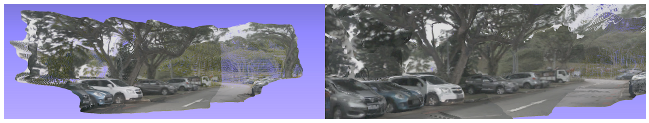


Figure 6: NVS with different frame length: 2(L), 8(R)

Privacy-Safety Trade-off: A primary concern with location obfuscation is that shared data from a synthesized pipeline may compromise fidelity. We address this trade-off via three perspectives.

- **Long-term reconstruction accuracy:** VGGT-Long [9] demonstrates that feed-forward 3D reconstruction can scale to kilometer-level outdoor environments while maintaining high fidelity. On the Waymo Open Dataset, VGGT-Long achieves an average Chamfer Distance of 2.021m—significantly outperforming traditional methods like DROID-SLAM (4.870m). Furthermore, it successfully reconstructs trajectories spanning up to 3.7 kilometers on KITTI sequences and maintains stable performance across diverse weather conditions, such as fog, rain, and sunset, in Virtual KITTI.
- **High Quality NVS:** Regarding novel view synthesis, 3DGS methods on EUVS benchmark [15] achieve a PSNR of 19.53 dB and SSIM of 0.75 for interpolated views. This quality can be further augmented by video diffusion-based approaches; for example, ReconDreamer++ [45] attains a PSNR of 20.52 dB, SSIM of 0.7033, and LPIPS of 0.3057. Crucially, these synthesized views preserve the geometric accuracy required for downstream perception: ReconDreamer++ achieves a Novel Trajectory Agent IoU of 0.365 and Lane IoU of 51.33 for 3-meter lateral shifts on Waymo. These results indicate that privacy-preserving NVS generates shared data of sufficient quality for cooperative perception.
- **Inherent robustness of cooperative perception:** Even if shared data contains inaccuracies, cooperative perception systems are inherently designed to handle erroneous inputs. Methods like ROBOSAC [23] introduce sampling-based defense strategies, comparing results from random subsets of teammates to reach consensus, which enables robustness against 80–90% outliers. Similarly, CP-Guard [17] employs probability-agnostic sample consensus (PASAC) to detect and eliminate inconsistent data without requiring prior knowledge of error distributions. These approaches demonstrate that collaborative perception naturally incorporates mechanisms to verify data consistency and reject outliers, regardless of whether they stem from adversarial attacks or reconstruction artifacts.

5 Open Questions & Discussion

Vision foundation models: While vision foundation models have revolutionized 3D understanding, key challenges remain: (1) extending robust performance from static to dynamic scenes, (2) obtaining metric-scale rather than relative depth estimates, (3) achieving occlusion-robust view selection, and (4) ensuring resilience against sophisticated adversarial attacks (e.g., physics-based material reflection analysis), (5) extending to other spatial sensors (radar, lidar). Despite these challenges, feedforward models like VGGT show promise for low-latency 3D perception.

Hierarchical data fidelity: Current cooperative perception systems operate with either all-processed or all-raw data. Given raw data's bandwidth demands, we advocate for hierarchical fidelity with opportunistic switching to maximize data quality while activating privacy preservation when sharing raw data. This requires mechanisms for activating, maintaining, and deactivating SHARP's privacy methods, with strategies to bootstrap open stack adoption.

Privacy as a first class citizen for networked vehicles: Raw data cooperative perception demands privacy-by-design alongside bandwidth, latency, synchronization, sensor interference, and trust considerations. While SHARP provides high-utility, privacy-safe solutions, it doesn't address all constraints. Networks may contain mixed flows, leading to orchestration and rate control challenges.

World privacy: While roads are public spaces, raw data sharing risks transforming vehicles into mass surveillance systems tracking pedestrians and license plates. Camera data particularly demands privacy safeguards—techniques like face and license plate anonymization before backend upload [35] should be incorporated into open stacks to balance raw data benefits with world privacy protection.

Implications for off-vehicle world understanding: Privacy safe raw data processing extends beyond real-time cooperative perception to edge and cloud services like high-resolution map building. While currently dominated by specialized fleets (e.g., Google Street View), SHARP enables democratized, crowd-sourced mapping accessible to diverse stakeholders (smart city planners, municipalities) at significantly lower cost and higher update frequency.

6 Conclusion

Multi-stakeholder groups are currently working to implement vehicle-to-vehicle communication for sharing highly-processed, high-priority safety messages [1, 25]. This paper looks even further ahead, exploring the exchange of high-fidelity, raw spatial sensor data between vehicles. For this to be practically adopted, privacy must be a primary design consideration. We've identified several privacy concerns that come with sharing raw data and proposed a comprehensive research agenda to address them. The goal of this paper is to initiate research and international discussions on this topic, aiming for the eventual adoption of raw data networking, similar to what's happening with processed data.

Acknowledgements

The authors are supported in part by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation and by the US National Science Foundation through grants: 2504963, 2312716, 2212688, 2112562, and 2107060.

References

[1] ITS America. 2023. ITS America National V2X Deployment Plan.
[2] D. Archer, B. Pigem, D. Bogdanov, et al. 2023. UN handbook on privacy-preserving computation techniques. *arXiv* (2023).
[3] K. Bansal, K. Rungta, S. Zhu, et al. 2020. Pointillism: Accurate 3d bounding box estimation with multi-radars. In *SenSys*.
[4] C. Bloom, J. Tan, et al. 2017. Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In *SOUPS*.
[5] H. Caesar, V. Bankiti, A. H. Lang, and othersRetry. 2020. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*.
[6] Z. Cai and A. Xiong. 2023. Understand Users' Privacy Perception and Decision of V2X Communication in Connected Autonomous Vehicles. In *USENIX Security*.
[7] Federal Trade Commission. 2024. Cars Consumer Data: On Unlawful Collection Use.
[8] R. Das, A. Gadre, S. Zhang, et al. 2018. A deep learning approach to IoT authentication. In *IEEE ICC*.
[9] K. Deng, Z. Ti, J. Xu, et al. [n. d.]. VGGT-Long: Chunk it, Loop it, Align it – Pushing VGGT's Limits on Kilometer-scale Long RGB Sequences.
[10] A. Dosovitskiy, G. Ros, F. Codevilla, et al. 2017. CARLA: An Open Urban Driving Simulator. In *CoRL*.
[11] M. Enev, A. Takakuwa, K. Koscher, et al. 2016. Automobile driver fingerprinting. *Proceedings on Privacy Enhancing Technologies* (2016).
[12] ETSI. 2024. Intelligent Transport System (ITS); Vehicular Communications; Basic Set of Applications; Collective Perception Service.
[13] Y. Fu, S. Liu, et al. 2024. Colmap-free 3d gaussian splatting. In *CVPR*.
[14] C. Godard, O. Mac Aodha, M. Firman, et al. 2019. Digging into self-supervised monocular depth estimation. In *ICCV*.
[15] X. Han, Z. Jia, B. Li, et al. [n. d.]. Extrapolated Urban View Synthesis Benchmark. *ICCV*.
[16] K. Henry. 2020. Pseudonym Issuing Strategies for Privacy-Preserving V2X Communication. *SAE International Journal of Transportation Cybersecurity and Privacy* 2, 11-02-02-0012 (2020).
[17] S. Hu, Y. Tao, G. Xu, et al. [n. d.]. CP-Guard: Malicious Agent Detection and Defense in Collaborative Bird's Eye View Perception. *AAAI* ([n. d.]).
[18] Y. Hu, S. Fang, Z. Lei, et al. 2022. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *NeurIPS*.
[19] N. Keetha, N. Müller, J. Schönberger, et al. 2025. MapAnything: Universal feed-forward metric 3D reconstruction. *arXiv* (2025).
[20] B. Kerbl, G. Kopanas, T. Leimkühler, et al. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023).
[21] K. Khan, A. Khalid, Y. Turkar, et al. 2024. VRF: Vehicle Road-side Point Cloud Fusion. In *ACM MobiSys*.
[22] KPMG. 2020. Automotive Data Sharing.
[23] Y. Li, Q. Fang, J. Bai, et al. [n. d.]. Among Us: Adversarially Robust Collaborative Perception by Consensus. *arXiv* ([n. d.]).
[24] Y. Liu, F. Wang, N. Wang, et al. 2023. Echoes beyond points: Unleashing the power of raw radar data in multi-modality fusion. *NeurIPS* (2023).
[25] Landline Media. 2024. U.S. DOT launches framework to deploy vehicle-to-everything technology nationwide.
[26] E. Mikhail, J. Bethel, and J. McGlone. 2001. *Introduction to modern photogrammetry*. John Wiley & Sons.
[27] B. Mildenhall, P. Srinivasan, M. Tancik, et al. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *CACM* (2021).
[28] T. Naseer, W. Burgard, and C. Stachniss. 2018. Robust visual localization across seasons. *IEEE Trans. Robot.* 34, 2 (2018).
[29] X. Ren, T. Shen, J. Huang, et al. 2025. GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control. In *CVPR*.
[30] J. Schönberger and J. Frahm. 2016. Structure-from-Motion Revisited. In *CVPR*.
[31] J. Shenoy, Z. Liu, B. Tao, et al. 2022. Rf-protect: privacy against device-free human tracking. In *SIGCOMM*.
[32] C. S. Shin, W. Pang, C. Li, et al. 2024. RECAP: 3D traffic reconstruction. In *ACM MobiCom*.
[33] S. Sumikura, M. Shibuya, and K. Sakurada. 2019. OpenVSLAM: A versatile visual SLAM framework. In *ACM MM*.
[34] Z. Teed and J. Deng. 2021. DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras. *NeurIPS* (2021).
[35] Tesla. 2025. Customer Privacy Notice.
[36] C. Toft, W. Maddern, A. Torii, et al. 2020. Long-term visual localization revisited. *IEEE TPAMI* 44, 4 (2020).
[37] J. Wang, M. Chen, N. Karaev, et al. 2025. Vggt: Visual geometry grounded transformer. In *CVPR*.
[38] S. Wang, V. Leroy, Y. Cabon, et al. 2024. Dust3r: Geometric 3d vision made easy. In *CVPR*.
[39] T. Wang, S. Manivasagam, M. Liang, et al. 2020. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *ECCV*.
[40] H. Runsheng Xu. 2022. OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with V2V Communication. In *ICRA*.
[41] R. Xu, H. Xiang, Z. Tu, et al. 2022. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *ECCV*. Springer.
[42] D. Yang, K. Yang, Y. Wang, et al. [n. d.]. How2comm: Communication-efficient and collaboration-pragmatic multi-agent perception. *NeurIPS*.
[43] K. Yang, D. Yang, et al. 2023. What2comm: Towards communication efficient collaborative perception via feature decoupling. In *ACM MM*.
[44] Q. Zhang, X. Zhang, R. Zhu, et al. 2023. Robust real-time multi-vehicle collaboration on asynchronous sensors. In *MobiCom*.
[45] G. Zhao, X. Wang, C. Ni, et al. [n. d.]. ReconDreamer++: Harmonizing Generative and Reconstructive Models for Driving Scene Representation. *arXiv*.