Evidence Alignment and Transfer Boundaries in Road-Scene Geometry

Abstract

A central challenge in road-scene geometry is to compare studies whose mechanisms and validation settings do not share a single denominator. The present review uses attention-based fusion of LiDAR and camera cues for curb detection and a wavelet-transform state-space decoder for zero-shot depth estimation as focal cases for a boundary-aware synthesis. Two target papers are triangulated against 12 locally validated publications. The comparison follows sensor fusion, depth priors, attention efficiency, domain transfer, road safety and deliberately separates mechanistic interpretation from performance ranking, because the latter can conceal incompatible experimental or operational conditions. Across the evidence base, the decisive issue is alignment: sensor fusion shapes what is observed, depth priors shapes how it is compared, and road safety governs whether the conclusion can be transferred. Uncertainty is most informative when reported as part of the result rather than treated as a postscript. The contribution is a decision-oriented synthesis that connects method selection to failure cost and treats reproducibility, provenance, and bounded generalization as first-order design requirements.

Keywords

Road-Scene Geometry; Sensor Fusion; Depth Priors; Attention Efficiency; Domain Transfer; Road Safety

1. Introduction

Inquiry into road-scene geometry must negotiate scientific ambition and practical constraint. Progress requires not only stronger nominal performance but also explanations that explain both success and failure. The tension becomes concrete in comparing attention-based fusion of LiDAR and camera cues for curb detection with a wavelet-transform state-space decoder for zero-shot depth estimation through evidence alignment and transfer boundaries. A result that is compelling in a particular material, corpus, cohort, geography, or load profile may be fragile outside its original boundary. The review process consequently has to preserve the unit of analysis and the decision context. Equally important is distinguishing a mechanism from a correlation, a benchmark advantage from a deployment advantage, and an interpretable diagnostic from a causal explanation.

The review adopts five complementary perspectives: sensor fusion, depth priors, attention efficiency, domain transfer, road safety. This set is useful because they jointly cover the technical object, measurement chain, and prerequisites of external validity. The review is anchored in representation, objective, and evaluation protocol. Under that principle, novel technique alone cannot settle the comparison; the comparison also requires aligned controls, explicit uncertainty, and credible resource or safety assumptions. The article is a literature synthesis and does not claim new experiments, participants, measurements, or performance results.

2. Methodological Foundations

The evidence pathway can be divided into the input evidence, the transformation applied to it, the output used for evaluation, and the decision that follows. In road-scene geometry, the chain is commonly fragmented even though errors propagate across them. Upstream noise may grow as it passes through a flexible model; an apparently accurate output can still be poorly calibrated for the final decision. As a consequence, ablation evidence should accompany aggregate performance. Sound comparative work specifies its controlled factors and permitted variation, then selects metrics that correspond to the intended use rather than to convenience alone.

Boundary awareness provides the next principle. Sensor fusion defines the local design problem, while road safety determines whether the design can survive outside that boundary. Between these endpoints, depth priors and attention efficiency connect those ends by exposing trade-offs that a single score conceals. This is where negative results and sensitivity evidence maps the conditions under which the explanation remains viable. For applied work, documentation of deployment constraints is therefore part of the scientific result, not an administrative afterthought.

3. Comparative Evaluation

The target study X2-01, Falcon: Fused attention for lidar-camera curb detection [1], addresses a concrete part of road-scene geometry through attention-based fusion of LiDAR and camera cues for curb detection. It identifies a representation, mechanism, or evaluation boundary needed to compare claims. Against the focus on comparing attention-based fusion of LiDAR and camera cues for curb detection with a wavelet-transform state-space decoder for zero-shot depth estimation through evidence alignment and transfer boundaries, the paper is read as a case whose boundary must be retained: transfer may depend on its sensing regime, preparation, workload, population, or benchmark. The synthesis records the design boundary before comparing assumptions across sources.

The target study X2-02, WaveSamba: A Wavelet Transform SSM Zero-Shot Depth Estimation Decoder [2], addresses a concrete part of road-scene geometry through a wavelet-transform state-space decoder for zero-shot depth estimation. It identifies a representation, mechanism, or evaluation boundary needed to compare claims. Against the focus on comparing attention-based fusion of LiDAR and camera cues for curb detection with a wavelet-transform state-space decoder for zero-shot depth estimation through evidence alignment and transfer boundaries, the paper is read as a case whose boundary must be retained: transfer may depend on its sensing regime, preparation, workload, population, or benchmark. The synthesis records the design boundary before comparing assumptions across sources.

Across the focal studies, sensor fusion should be interpreted jointly with depth priors. A design can improve one but transfer cost into the coupled dimension. The analysis can be summarized in a compact matrix whose rows are operating conditions and whose columns are evidence quality, computation or process cost, robustness, and consequence of failure. The matrix is designed to prevent an attractive mechanism from being detached from the circumstances that support it. The same device identifies which ablations are informative: removing a component should test a stated causal role, not merely create another number.

Attention efficiency translates method behavior into decision evidence. The evaluation protocol needs to separate familiar inputs from perturbations and retain distributional summaries, and test plausible alternative explanations. Where distribution shift is central, calibration or sensitivity is as important as ranking. Where costs or hazards are asymmetric, utility-weighted assessment is more revealing than undifferentiated accuracy. These choices do not erase study differences; they make the reasons for their differences auditable.

4. Architecture and Learning Objectives

A complementary strand is visible in Robust Curb Detection with Fusion of 3D-Lidar and Camera Data [3]; MF-BEVFusion: multiscale depth estimation and fully dynamic fusion for camera-LiDAR BEV 3D object detect... [4]; Dense Depth Map Estimation Based on Camera–LiDAR Sensor Fusion [5]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of road-scene geometry. Together they illuminate sensor fusion: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

For triangulation, the literature includes Zero-Shot Polarization-Intensity Physical Fusion Monocular Depth Estimation for High Dynamic Range Scenes [6]; DepthFusion: Depth-Aware Hybrid Feature Fusion for LiDAR-Camera 3D Object Detection [7]; Probabilistic multi-modal depth estimation based on camera–LiDAR sensor fusion [8]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of road-scene geometry. Together they illuminate depth priors: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

The design space is broadened by Obstacle detection for intelligent robots based on the fusion of 2D lidar and depth camera [9]; Real-Time Object Detection and Distance Measurement Enhanced with Semantic 3D Depth Sensing Using Camera... [10]; A LiDAR-depth camera information fusion method for human robot collaboration environment [11]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of road-scene geometry. Together they illuminate attention efficiency: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

A first comparison set connects Obstacle detection for intelligent robots based on the fusion of 2D lidar and depth camera [12]; Adaptive Active Fusion of Camera and Single-Point LiDAR for Depth Estimation [13]; Sparse LiDAR and Stereo Fusion (SLS-Fusion) for Depth Estimation and 3D Object Detection [14]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of road-scene geometry. Together they illuminate domain transfer: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

5. Deployment and Governance

For implementation, the review suggests a staged workflow. First, define the decision and its failure cost. Second, specify the evidence and operating envelope. Third, choose a method whose inductive assumptions match that evidence. Fourth, test domain transfer and road safety under controlled perturbations. Finally, retain provenance so that an unexpected outcome can be traced to data, configuration, material batch, environmental condition, or user interaction. The staged design keeps comparing attention-based fusion of LiDAR and camera cues for curb detection with a wavelet-transform state-space decoder for zero-shot depth estimation through evidence alignment and transfer boundaries connected to observable requirements.

Across applications, responsible progress in road-scene geometry rests on interactions among components. Interfaces determine how uncertainty moves between measurement and inference, between algorithm and operator, or between microscopic design and system behavior. A shared protocol should state acceptance conditions and what would overturn a claim. When those agreements are explicit, optimization need not trade away validity, safety, or intelligibility without notice.

6. Limitations and Future Directions

Interpretation is bounded by cross-study variation in labels, design choices, and reporting precision. A valid DOI record authenticates bibliographic facts without independently certifying empirical conclusions. Versioning is another source of uncertainty. The focal citations supplied as Scholar-verified records were preserved; uncertain or malformed records were documented separately rather than silently rewritten.

A productive research agenda would preregister comparison protocols where feasible, retain machine-readable setup details while testing at least one nontrivial shift in deployment constraints. Research should also classify failures by causal mechanism as well as prevalence. For road-scene geometry, a valuable shared benchmark would combine standard-condition utility, efficiency, confidence quality, and fault recovery. Such a benchmark would reward designs that remain useful when evidence is incomplete or conditions change.

7. Conclusion

The literature on road-scene geometry is better interpreted through the relationships among studies rather than a collection of isolated scores. The focal studies show how comparing attention-based fusion of LiDAR and camera cues for curb detection with a wavelet-transform state-space decoder for zero-shot depth estimation through evidence alignment and transfer boundaries; the additional literature reveals the assumptions required to interpret those contributions. Across sensor fusion, depth priors, attention efficiency, domain transfer, and road safety, a durable conclusion is the need for alignment: the representation or mechanism must match the problem, the metric must serve the decision while deployment remains inside the tested boundary. The consequence is evidence that travels more safely and fails more informatively.

References

1. Chen, Y., Long, Z., Wu, Y., & Chen, L. (2024). Falcon: Fused attention for lidar-camera curb detection.

2. Hou, B., & Long, Z. (2024). WaveSamba: A Wavelet Transform SSM Zero-Shot Depth Estimation Decoder. In 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 738-744). IEEE.

3. Tan, J., Li, J., An, X., & He, H. (2014). Robust Curb Detection with Fusion of 3D-Lidar and Camera Data. Sensors, 14(5), 9046-9073. https://doi.org/10.3390/s140509046
4. Liu, J., Yue, S., Hao, W., & Cai, Y. (2026). MF-BEVFusion: multiscale depth estimation and fully dynamic fusion for camera-LiDAR BEV 3D object detection. Journal of Electronic Imaging, 35(02). https://doi.org/10.1117/1.jei.35.2.023009
5. Bong, E. J., & Kee, S. C. (2026). Dense Depth Map Estimation Based on Camera–LiDAR Sensor Fusion. IEEE Sensors Journal, 26(4), 5891-5901. https://doi.org/10.1109/jsen.2025.3649237
6. Rao, R., Ouyang, Z., Chen, S., Chen, L., Huang, G., & Cui, C. (2026). Zero-Shot Polarization-Intensity Physical Fusion Monocular Depth Estimation for High Dynamic Range Scenes. Photonics, 13(3), 268. https://doi.org/10.3390/photonics13030268
7. Ji, M., Yang, J., & Zhang, S. (2026). DepthFusion: Depth-Aware Hybrid Feature Fusion for LiDAR-Camera 3D Object Detection. IEEE Transactions on Multimedia, 28, 7217-7227. https://doi.org/10.1109/tmm.2026.3668596
8. Obando-Ceron, J. S., Romero-Cano, V., & Monteiro, S. (2023). Probabilistic multi-modal depth estimation based on camera–LiDAR sensor fusion. Machine Vision and Applications, 34(5). https://doi.org/10.1007/s00138-023-01426-x
9. Meng, L., Zhao, H., & Fan, B. (2024). Obstacle detection for intelligent robots based on the fusion of 2D lidar and depth camera. International Journal of Hydromechatronics, 7(1). https://doi.org/10.1504/ijhm.2024.10060856
10. Yildiz, A. S., Meng, H., & Swash, M. R. (2025). Real-Time Object Detection and Distance Measurement Enhanced with Semantic 3D Depth Sensing Using Camera–LiDAR Fusion. Applied Sciences, 15(10), 5543. https://doi.org/10.3390/app15105543
11. Wang, Z., Li, P., Zhang, Q., Zhu, L., & Tian, W. (2025). A LiDAR-depth camera information fusion method for human robot collaboration environment. Information Fusion, 114, 102717. https://doi.org/10.1016/j.inffus.2024.102717
12. Fan, B., Zhao, H., & Meng, L. (2024). Obstacle detection for intelligent robots based on the fusion of 2D lidar and depth camera. International Journal of Hydromechatronics, 7(1), 67-88. https://doi.org/10.1504/ijhm.2024.135994
13. Tran, D. M., Ahlgren, N., Depcik, C., & He, H. (2023). Adaptive Active Fusion of Camera and Single-Point LiDAR for Depth Estimation. IEEE Transactions on Instrumentation and Measurement, 72, 1-9. https://doi.org/10.1109/tim.2023.3284129
14. Mai, N. A. M., Duthon, P., Khoudour, L., Crouzil, A., & Velastin, S. A. (2021). Sparse LiDAR and Stereo Fusion (SLS-Fusion) for Depth Estimation and 3D Object Detection. IET Conference Proceedings, 2021(1), 150-156. https://doi.org/10.1049/icp.2021.1442