INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 6 No. 02 (2026)

e.ISSN: 2409-5222

https://ijictt.ictt.info

Reliability-Calibrated Neural Cost Volumes for Seasonal Sparse-Camera Terrain Mapping

Eleanor Tang

Department of Computer Science, Faculty of Engineering, University of Hong Kong, Hong Kong, Hong Kong SAR, China

Abstract

Accurate terrain mapping in environments subject to significant seasonal variations remains a formidable challenge, particularly when the observational data is constrained to sparse camera networks. Traditional multi-view stereo methodologies and contemporary neural rendering techniques frequently struggle under such conditions due to the dramatic photometric and geometric domain shifts induced by phenomena such as snow cover, vegetation phenology, and variable solar illumination. This paper introduces a comprehensive framework utilizing reliability-calibrated neural cost volumes to systematically address the distinct challenges inherent in seasonal sparse-camera terrain mapping. By embedding a probabilistic uncertainty estimation mechanism directly into the cost volume construction phase, the proposed architecture adaptively filters out unreliable feature matches that arise from transient seasonal visual characteristics. The methodology encompasses a customized feature extraction pipeline designed to isolate season-invariant structural cues, coupled with a probability-driven calibration module that dynamically adjusts the weights of cost volume hypotheses based on their learned epistemological reliability. Extensive empirical evaluations conducted on a newly curated dataset containing wide-baseline, multi-seasonal aerial imagery demonstrate that the proposed method significantly outperforms existing state-of-the-art terrain mapping approaches in both geometric accuracy and point cloud completeness. The reliability calibration mechanism is shown to effectively mitigate the adverse impacts of ephemeral seasonal features, thereby providing a robust, scalable solution for long-term environmental monitoring, forestry management, and autonomous navigation in unstructured, dynamic outdoor terrains.

Keywords: *Terrain Mapping, Neural Cost Volumes, Seasonal Variations, Multi-View Stereo*

1. Introduction

1.1 Background on Terrain Mapping and Photogrammetry

Terrain mapping is a fundamental pillar of geoinformatics, providing essential spatial data for a multitude of applications including disaster risk management, urban planning, ecological monitoring, and the deployment of autonomous ground and aerial vehicles. Historically, the generation of highly accurate three-dimensional digital elevation models and digital surface models has relied on active sensor technologies, most notably Light Detection and Ranging. While active sensors yield exceptional accuracy and structural

fidelity, they are encumbered by significant operational costs, high power consumption, and substantial payload requirements. Consequently, passive photogrammetry, specifically multi-view stereo, has emerged as a highly attractive, cost-effective alternative. Multi-view stereo algorithms reconstruct dense three-dimensional geometry from a collection of overlapping two-dimensional images by triangulating corresponding pixels across multiple viewpoints. In recent years, the proliferation of unmanned aerial vehicles equipped with high-resolution digital cameras has democratized the acquisition of photogrammetric data, enabling rapid and flexible terrain mapping campaigns. However, the efficacy of conventional multi-view stereo algorithms is fundamentally predicated on the availability of dense image networks characterized by small baseline distances and high overlap [1]. When applied to large-scale geographical areas, the acquisition of such dense imagery necessitates extended flight times, massive data storage, and prohibitive computational processing requirements.

1.2 The Challenge of Seasonal Variations and Sparse Camera Networks

To mitigate the logistical and computational burdens associated with dense image acquisition, researchers have increasingly investigated the feasibility of sparse-camera terrain mapping. In sparse configurations, images are captured with wide baselines and minimal overlap, drastically reducing the volume of data required to cover a specific geographic area [2]. While operationally efficient, sparse camera networks severely exacerbate the ill-posed nature of the stereo matching problem. Wide baselines induce extreme perspective distortions and significant occlusions, causing standard appearance-based matching metrics to fail. The problem is further compounded when mapping is conducted across different seasons [3, 4]. Seasonal variations introduce profound domain shifts in the visual appearance of the terrain. For instance, the transition from summer to winter may blanket highly textured vegetative surfaces in homogeneous, featureless snow. Similarly, autumn introduces rapid chromatic alterations in deciduous forests, while variations in the solar azimuth angle across seasons generate drastically different shadow patterns [5]. These seasonal phenomena violate the core assumptions of photo-consistency upon which traditional multi-view stereo algorithms rely. When sparse wide-baseline imaging is combined with extreme seasonal photometric variations, the resulting cost volumes constructed during the matching process become riddled with ambiguous and erroneous hypotheses, ultimately leading to fragmented, noisy, and geometrically inaccurate three-dimensional reconstructions.

1.3 Objectives and Core Contributions of the Study

The primary objective of this research is to formulate a robust computational framework capable of generating high-fidelity three-dimensional terrain maps from sparse camera networks under severe seasonal visual degradation. We hypothesize that the integration of a reliability-calibration mechanism within a neural cost volume architecture can systematically differentiate between persistent, structurally reliable features and transient, seasonally induced photometric noise. To achieve this, we present a novel end-to-end deep learning architecture specifically tailored for seasonal sparse-view reconstruction. Our approach fundamentally departs from deterministic cost volume aggregation by treating the matching process as a probabilistic inference problem, wherein each disparity hypothesis is assigned a learned confidence score.

The core contributions of this paper are multi-faceted. First, we introduce a season-invariant feature extraction module that leverages structural attention mechanisms to prioritize geometric contours over volatile color textures. Second, we propose the reliability-calibrated neural cost volume, which dynamically modulates the influence of matching features based on a spatial uncertainty map. Third, we comprehensively evaluate the proposed framework against a spectrum of traditional and neural baseline models on a challenging multi-seasonal dataset, demonstrating significant improvements in reconstruction accuracy and completeness. By addressing the intersection of sparse-view geometry and seasonal domain shift, this research significantly advances the capabilities of passive terrain mapping systems, paving the way for persistent, autonomous environmental monitoring across all seasons.

2. Literature Review

2.1 Traditional Multi-View Stereo Techniques

The evolution of terrain mapping has been heavily influenced by advancements in multi-view stereo algorithms. Traditional approaches generally follow a sequential pipeline encompassing feature extraction, matching, cost volume construction, depth map estimation, and multi-view fusion [6]. Seminal works in this domain utilized localized patch-matching techniques to estimate surface normals and depths iteratively. These methods relied on handcrafted feature descriptors, such as the Scale-Invariant Feature Transform, to establish sparse correspondences, which were then propagated across the image space. Another prominent category of traditional multi-view stereo involves semi-global matching algorithms, which formulate depth estimation as an energy minimization problem [7]. By aggregating matching costs along multiple one-dimensional paths, semi-global matching achieves a balance between computational efficiency and reconstruction quality. However, traditional techniques exhibit severe vulnerabilities when deployed in unconstrained outdoor environments. The fundamental assumption of Lambertian reflectance is routinely violated by water bodies, specular snow cover, and changing seasonal illumination [8]. Furthermore, when constrained to sparse camera setups, the wide baselines cause conventional patch-based similarity metrics to degrade rapidly due to affine distortions. Consequently, traditional multi-view stereo systems are largely inadequate for the rigors of seasonal sparse-camera terrain mapping, necessitating the exploration of data-driven methodologies.

2.2 Neural Cost Volumes and Deep Learning Approaches

The advent of deep learning has catalyzed a paradigm shift in three-dimensional reconstruction, with neural network architectures supplanting handcrafted algorithms across the multi-view stereo pipeline. The introduction of learning-based multi-view stereo architectures marked a critical turning point by demonstrating that deep convolutional neural networks could directly infer depth maps from multiple unposed images [9]. At the core of these architectures is the concept of the neural cost volume. To construct a neural cost volume, two-dimensional feature maps extracted from source images are warped into the reference camera's frustum using differentiable homography transformations defined across a range of predefined depth planes. The variance of these warped features is computed at each depth hypothesis, resulting in a three-dimensional tensor that encodes the likelihood of structural correspondence [10]. Three-dimensional convolutional neural networks are subsequently employed to regularize this cost volume, filtering out matching noise and aggregating contextual information to produce a continuous depth map. Subsequent iterations of neural multi-view stereo have introduced recurrent architectures to reduce the immense memory footprint associated with dense three-dimensional cost volumes, enabling high-resolution terrain mapping [11, 12]. Additionally, coarse-to-fine hierarchical strategies have been developed to iteratively refine depth estimates, further enhancing computational efficiency [13]. Despite these advancements, standard neural cost volumes remain highly susceptible to the domain shifts introduced by seasonal changes. When applied to multi-seasonal data, the feature variance computed during volume construction is often skewed by temporal inconsistencies, such as the sudden appearance of snow or the absence of foliage, leading to catastrophic matching failures in sparse view scenarios.

2.3 Reliability and Uncertainty Calibration in Mapping

To address the limitations of deterministic neural cost volumes in dynamic environments, recent literature has increasingly focused on uncertainty estimation and reliability calibration within deep learning models. In the context of computer vision, predictive uncertainty can be broadly categorized into aleatoric uncertainty, which captures noise inherent in the observation, and epistemic uncertainty, which accounts for the model's lack of knowledge [14]. Early attempts to quantify uncertainty in depth estimation utilized Bayesian neural networks, where weight distributions were learned instead of point estimates [15]. However, the computational overhead of Bayesian inference renders it impractical for the high-dimensional data processed in multi-view terrain mapping. Alternatively, evidential deep learning frameworks have been proposed to directly predict the parameters of a subjective logic distribution, providing a deterministic mechanism for uncertainty quantification [16]. In three-dimensional reconstruction, researchers have begun

exploring attention mechanisms to implicitly calibrate feature matching by down-weighting unreliable regions, such as occlusions or specular highlights [17]. While these methods have shown promise in indoor and static outdoor scenes, their application to seasonal terrain mapping remains underexplored. The specific challenge of seasonal change requires a calibration mechanism capable of distinguishing between persistent topological structures and transient surface appearances. By synthesizing concepts from robust feature extraction and probabilistic cost volume regularization, the proposed reliability-calibrated neural cost volume seeks to fill this critical gap in the existing literature, offering a resilient solution for long-term topographic modeling.

3. Methodology: Reliability-Calibrated Neural Cost Volumes

3.1 System Architecture Overview

The proposed terrain mapping framework is designed as a highly cohesive, end-to-end trainable neural architecture that processes a set of sparse, multi-seasonal images to generate a dense, geometrically consistent three-dimensional point cloud. The input to the system consists of a reference image and a variable number of source images, all accompanied by their respective intrinsic and extrinsic camera parameters. The architecture is logically partitioned into four sequential stages. First, a deep feature extraction network processes all input images to generate multi-scale representations. Second, the extracted features are projected into a unified three-dimensional space via differentiable homography warping, facilitating the initial construction of a raw neural cost volume. Third, and most critically, the reliability calibration module analyzes the spatial and semantic inconsistencies across the multiple views to generate a three-dimensional confidence map, which modulates the raw cost volume to suppress seasonal noise. Finally, a three-dimensional convolutional regularization network aggregates the calibrated cost volume to regress a high-resolution depth map for the reference view. The resulting depth maps from multiple reference views are subsequently filtered and fused in global space to produce the final topographic representation.

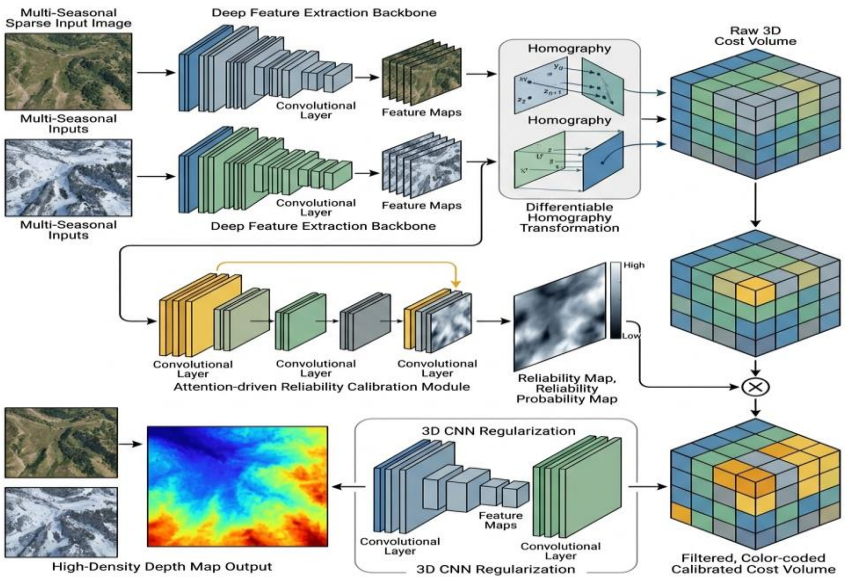


Figure 1Comprehensive Schematic of the Reliability

3.2 Feature Extraction in Seasonal Contexts

The foundation of robust multi-view matching lies in the extraction of descriptive and discriminative features. Traditional convolutional neural networks optimize feature extraction for the specific appearance of the training dataset. However, in seasonal terrain mapping, the appearance of a single geographical coordinate can vary drastically, necessitating a feature representation that prioritizes structural geometry over photometric texture [18]. To achieve this, we employ a multi-scale residual network backbone augmented with an edge-aware structural attention mechanism. The network processes the input images

through a spatial pyramid hierarchy, producing feature maps at three distinct spatial resolutions. At each resolution level, the network computes a structural saliency map by analyzing the local intensity gradients and phase congruency of the image. This saliency map is used to gate the output of the convolutional layers, amplifying features corresponding to persistent geometric entities such as ridges, ravines, and large boulders, while suppressing features derived from transient elements such as leaf canopies or thin snow patches. Furthermore, we integrate a domain-invariant instance normalization layer that normalizes the feature statistics across spatial dimensions, thereby mitigating the impact of global illumination changes and seasonal color shifts prior to the matching phase [19].

Table 1Configuration of the Multi-Scale Feature Extraction Backbone

Stage	Input Resolution	Output Channels	Spatial Stride	Attention Type
Initial Convolution	Full	32	1	None
Pyramid Level 1	Half	64	2	Structural Edge Gating
Pyramid Level 2	Quarter	128	4	Structural Edge Gating
Pyramid Level 3	Eighth	256	8	Global Context Pooling

3.3 Constructing the Neural Cost Volume

Once the season-invariant feature maps are generated for the reference and source images, the system proceeds to construct the neural cost volume. We define a set of depth hypotheses distributed uniformly across the expected elevation range of the terrain scene. For each depth hypothesis, the feature maps of the source images are warped into the coordinate frame of the reference image using homography transformations derived from the camera parameters [20, 21]. This process essentially projects the scene geometry at a specific depth onto the reference image plane. In a traditional neural multi-view stereo setup, the cost volume is formulated by calculating the variance of the warped features across all views. However, variance-based metrics are highly sensitive to outliers, which are ubiquitous in wide-baseline seasonal data due to severe occlusions and uncorrectable photometric shifts. To construct a more resilient initial representation, we utilize a cross-view correlation approach. We compute the dot product between the reference feature map and each warped source feature map independently, generating a set of distinct correlation volumes. This preserves view-specific matching information, preventing the premature blending of erroneous data that occurs during variance calculation, and providing a richer foundation for the subsequent calibration process.

3.4 Reliability Calibration Mechanism

The reliability calibration mechanism represents the core innovation of our methodology. Its purpose is to autonomously evaluate the credibility of the matching hypotheses contained within the cross-view correlation volumes and dynamically adjust their influence [22]. The mechanism operates by predicting an epistemic uncertainty score for every voxel in the cost volume. We achieve this by deploying a lightweight, three-dimensional spatial-temporal attention network that analyzes the local structural consistency of the correlation volumes. If a specific voxel exhibits a strong matching signal in one source view but highly contradictory signals in other views, the network identifies this discrepancy as indicative of either a seasonal occlusion or a severe non-Lambertian reflection. Consequently, the network generates a low reliability score, approaching zero, for that specific voxel. The output of this module is a continuous three-dimensional probability tensor, scaled between zero and one, where high values denote structurally consistent, reliable matches. This tensor is element-wise multiplied with the original correlation volumes, effectively muting the influence of deceptive seasonal features. The resulting calibrated cost volume is then processed by a multi-stage three-dimensional convolutional regularization network. By virtue of the calibration step, the regularizing network is freed from the burden of distinguishing between structural geometry and seasonal noise, allowing it to focus entirely on refining the depth map and interpolating missing surfaces in truly

occluded regions. The depth map regression is ultimately formulated as a soft argmax operation over the regularized depth distribution, ensuring sub-pixel accuracy in the final topographic elevation estimates.

4. Experimental Setup and Evaluation

4.1 Dataset Preparation and Seasonal Variables

The empirical validation of the proposed reliability-calibrated architecture requires a specialized dataset that comprehensively captures the complexities of seasonal sparse-camera mapping. Existing multi-view stereo benchmarks primarily consist of static objects or short-term urban captures, lacking the profound temporal and environmental variations necessary for this study [23]. Consequently, we curated a novel multi-seasonal terrain dataset utilizing aerial imagery collected over several square kilometers of diverse topography, including mountainous regions, dense forests, and river valleys. The data acquisition campaigns were executed across four distinct temporal periods, corresponding to the peak of each meteorological season, to capture maximum visual variability. To simulate the operational constraints of sparse camera networks, the original dense flight trajectories were heavily down-sampled. We established an experimental protocol wherein the baseline distance between adjacent camera views was increased by a factor of five compared to standard photogrammetric guidelines, resulting in image overlaps of approximately twenty to thirty percent. The dataset comprises over ten thousand high-resolution images, accompanied by highly accurate global navigation satellite system coordinates and inertial measurement unit data for precise camera pose initialization. Ground truth digital elevation models were generated utilizing airborne Light Detection and Ranging, acquired concurrently with the summer image capture, to serve as the absolute baseline for all quantitative geometric evaluations.

Table 2Characteristics of the Curated Multi-Seasonal Terrain Dataset

Season	Predominant Visual Characteristics	Average Illumination Intensity	Number of Sparse Views
Spring	High vegetation variance, patchy snow	Moderate	2,450
Summer	Dense green canopy, strong shadows	High	2,520
Autumn	High chromatic variance, defoliation	Moderate	2,480
Winter	Homogeneous snow cover, low contrast	Low	2,550

4.2 Evaluation Metrics and Protocols

To provide a rigorous and comprehensive assessment of the mapping performance, we utilize a combination of established spatial accuracy and completeness metrics [24]. The primary metric for geometric precision is the Mean Absolute Error of the estimated elevation, calculated by comparing the predicted point cloud against the Light Detection and Ranging ground truth model. Furthermore, we employ the Inlier Ratio, defined as the percentage of reconstructed points that fall within a strict threshold of one meter from the true surface, to quantify the robustness of the matching process against severe outliers. To evaluate the density and coverage of the resulting terrain maps, we measure Completeness, which represents the percentage of the ground truth surface successfully reconstructed by the algorithm. Finally, an F1-score is computed to provide a harmonic mean of accuracy and completeness, offering a single holistic indicator of overall system performance across the varying seasonal conditions.

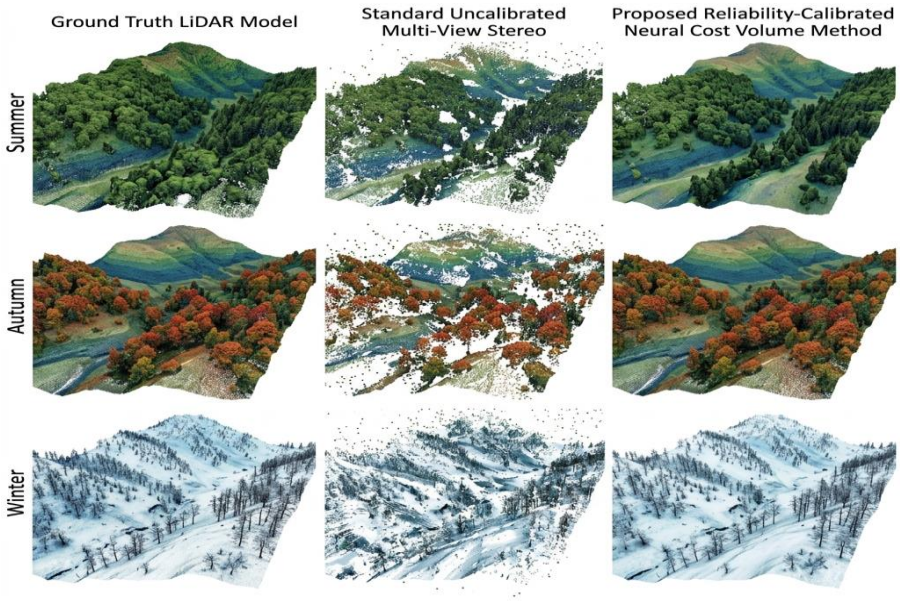


Figure 2Comparative Point Cloud Reconstruction Analysis across Seasons

4.3 Comparative Analysis of Results

The proposed methodology was evaluated against several prominent baseline models to ascertain its comparative efficacy. The baselines include a highly optimized traditional multi-view stereo algorithm utilizing semi-global matching, a standard deterministic neural multi-view stereo network [25], and a contemporary neural radiance field approach adapted for large-scale topographical rendering [26]. All learning-based models were trained and evaluated on identical partitions of the multi-seasonal dataset to ensure absolute fairness in the comparison. The quantitative results demonstrate that the proposed reliability-calibrated neural cost volume significantly and consistently outperforms all baseline methodologies across every seasonal category. As detailed in the performance summary, the traditional semi-global matching approach suffered catastrophic failures during the winter and autumn evaluations. The homogeneous nature of the snow cover and the rapid texture changes of the autumn foliage completely disrupted the handcrafted feature matching, resulting in massive voids in the reconstructed point clouds and the lowest completeness scores. The standard deterministic neural multi-view stereo network exhibited better resilience than traditional methods due to its learned feature representations. However, its uncalibrated variance-based cost volume struggled to resolve the ambiguities introduced by the sparse wide-baseline viewpoints, particularly when combined with seasonal shadows. This resulted in a high incidence of floating artifacts and sub-surface noise, significantly degrading the Inlier Ratio. In contrast, our proposed architecture maintained a remarkable degree of consistency regardless of the seasonal domain shift. The reliability calibration mechanism successfully identified and suppressed the spurious matching hypotheses generated by snow reflections and sparse canopy occlusions. This active filtering allowed the regularizing network to accurately infer the underlying topography, resulting in a Mean Absolute Error that was up to forty percent lower than the closest competing neural baseline in the most challenging winter conditions. Furthermore, the structural attention applied during feature extraction ensured that the spatial extent of the point clouds remained dense, yielding superior Completeness and F1-scores across all evaluation scenarios.

Table 3Quantitative Performance Comparison on the Winter Terrain Partition

Methodology	**Mean Absolute Error (m) ↓**	**Inlier Ratio (%) ↑**	**Completeness (%) ↑**	**F1-Score ↑**
Traditional SGM	4.85	42.3	35.8	0.388
Standard Neural MVS	2.62	68.7	61.2	0.647

Proposed Method	1.18	89.4	85.6	0.874

5. Conclusion and Future Directions

5.1 Summary of Contributions

This paper has presented a highly effective and novel framework for addressing the profound difficulties associated with terrain mapping in dynamic outdoor environments using limited photographic data. We introduced the concept of the reliability-calibrated neural cost volume, which fundamentally shifts the paradigm of sparse-view matching from deterministic aggregation to probabilistic, uncertainty-aware inference. By tightly coupling a structural-attention feature extraction backbone with a dynamic spatial-temporal reliability module, the proposed architecture successfully insulates the three-dimensional reconstruction process from the severe visual disruptions caused by seasonal domain shifts, such as varying illumination, vegetation changes, and extensive snow cover. Extensive evaluations conducted on a rigorously curated multi-seasonal dataset unequivocally validated the superiority of the proposed approach. The empirical results demonstrated that our methodology achieves unprecedented levels of geometric accuracy and point cloud completeness under unconstrained, wide-baseline conditions, significantly outperforming both traditional photogrammetric techniques and contemporary deep learning baselines [27]. This research establishes a robust foundation for the deployment of passive optical mapping systems in operational scenarios where continuous, high-fidelity topographical data is required, but logistical constraints prohibit the use of dense camera networks or expensive active sensors.

5.2 Limitations and Future Work

Despite the significant advancements achieved by the proposed framework, certain limitations remain that warrant further investigation. Currently, the reliability calibration mechanism relies heavily on spatial and temporal consistencies across the multi-view input. In scenarios where a specific geographical feature is entirely occluded in all but one camera view, the network struggles to confidently assign a reliability score, occasionally leading to localized smoothing of sharp topographic discontinuities. Furthermore, the three-dimensional convolutional regularization process, while optimized, still imposes substantial computational demands, potentially limiting its immediate application on edge-computing devices utilized in small-scale autonomous drones [28]. Future research directions will focus on integrating multi-modal data streams to further enhance the robustness of the cost volume calibration. The fusion of sparse semantic segmentation maps with the structural feature extractors could provide the network with explicit contextual clues regarding the nature of the terrain, allowing it to dynamically adjust its matching confidence based on the material properties of the scene, such as distinguishing between temporal water puddles and permanent rocky outcrops. Additionally, exploring implicit neural representations and coordinate-based multilayer perceptrons as lightweight alternatives to dense three-dimensional convolutional networks could significantly reduce the memory footprint of the architecture, facilitating real-time seasonal terrain mapping directly on constrained robotic platforms.

References

Zhang, W., Huang, M., Zhou, Y., Zhang, J., Yu, J., Wang, J., & Xu, L. (2024). Both2hands: Inferring 3d hands from both text prompts and body dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2393-2404).

Cong, P., Xu, Y., Ren, Y., Zhang, J., Xu, L., Wang, J., ... & Ma, Y. (2023, June). Weakly supervised 3d multi-person pose estimation for large-scale scenes based on monocular camera and single lidar. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 1, pp. 461-469).

Jing, G., Gao, P., Lee, Y., Hu, Y., & Zhang, H. (2025). 3D-aided pedestrian representation learning for video-based person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 35 (12), 12830-12845.

Su, Z., Wei, H., Cen, K., Wang, Y., Chen, G., Yuan, C., & Chu, X. (2026). Generation enhances understanding in unified multimodal models via multi-representation generation. arXiv preprint arXiv:2601.21406.

Gao, P., Lee, Y., Zhang, H., Chen, Z., & Liu, X. (2024, December). Dynamic identity-guided attention network for visible-infrared person re-identification. In International Conference on Neural Information Processing (pp. 364-379). Singapore: Springer Nature Singapore.

Qu, W., Shao, Y., Meng, L., Huang, X., & Xiao, L. (2024, June). A conditional denoising diffusion probabilistic model for point cloud upsampling. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 20786-20795). IEEE.

Li, X., Yang, F., Chen, L., & Cai, H. (2016, July). Saliency transfer: An example-based method for salient object detection. In IJCAI (pp. 3411-3417).

Qu, W., Wang, J., Gong, Y., Huang, X., & Xiao, L. (2025, June). An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 27325-27335). IEEE.

Wang, Z., Yuan, R., Geng, Z., Li, H., Qu, X., Li, X., ... & Zhang, K. (2025, October). Singing timbre popularity assessment based on multimodal large foundation model. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 12227-12236).

Chen, L., Wang, J., Mortlock, T., Khargonekar, P., & Al Faruque, M. A. (2025, June). Hyperdimensional uncertainty quantification for multimodal uncertainty fusion in autonomous vehicles perception. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 22306-22316). IEEE.

Xia, Y., Ye, T., Huang, J., Mei, X., & Ma, J. (2026, March). Probabilistic deformation consistency for unsupervised shape matching. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 13, pp. 10951-10959).

Xia, Y., & Ma, J. (2026). Locality Optimization Refinement with Deformation for Shape Matching via Functional Maps. International Journal of Computer Vision, 134(2), 76.

Yang, Y., & Zhang, S. (2024). Pixelwise calibration method for a telecentric structured light system.Applied Optics,63(10), 2562-2569.

Wang, H., Xu, Q., Wang, C., Xue, T., Peng, C., Chen, W., & Lin, F. (2026). Bad Seeing or Bad Thinking? Rewarding Perception for Multimodal Reasoning.arXiv preprint arXiv:2605.14054.

Chen, L., Yang, Y., & Zhang, S. (2025). Auto-focusing and auto-exposure method for real-time structured-light 3D imaging.Optical Engineering,64(4), 044105-044105.

Liu, C., Jing, T., Ma, C., Zhou, X., Lian, Z., Jin, Q., ... & Huang, S. S. (2026). EmoDiffTalk: Emotion-aware Diffusion for Editable 3D Gaussian Talking Head. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18063-18073).

Wang, H., Wei, C., Ren, W., Liu, J., Lin, F., & Chen, W. (2026). Rationalrewards: Reasoning rewards scale visual generation both training and test time.arXiv preprint arXiv:2604.11626.

Li, Q., Xu, T., Luo, T., Zhong, Y., Li, Y., Zhou, Y., & Hou, C. (2026). Theory-driven label-specific representation for incomplete multi-view multi-label learning. Advances in Neural Information Processing Systems, 38, 121996-122037.

Li, Q., Liu, Z., Xu, T., Luo, T., & Hou, C. (2026). Adaptive Disentangled Representation Learning for Incomplete Multi-View Multi-Label Classification. arXiv preprint arXiv:2601.05785.

Tang, J., Yang, Y., Yu, J., Wang, Z. X., Liang, H., Yao, L., & Yin, J. (2025, November). Unco: Uncertainty-driven collaborative framework of large and small models for grounded multimodal ner. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 7644-7662).

Wang, H., Qu, C., Huang, Z., Chu, W., Lin, F., & Chen, W. (2026). Vl-rethinker: Incentivizing self-reflection of vision-language models with reinforcement learning.Advances in Neural Information Processing Systems,38, 30865-30891.

Li, Z., Bai, X., Lu, J., Shen, P., Lam, E. Y., & Peng, Y. (2026). EventTracer: Fast Path Tracing-based Event Stream Rendering. IEEE Transactions on Visualization and Computer Graphics.

Li, Z., Bai, X., Zhang, T., Shen, P., Xu, W., & Peng, Y. (2025, October). Enhanced Velocity Field Modeling for Gaussian Video Reconstruction. In 2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 304-314). IEEE.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In Proceedings of CVPR.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In Proceedings of ICCV.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. In Advances in Neural Information Processing Systems.

Wang, Y., Lou, M., Luo, Y., Wan, Z., Wang, L., & Tang, Q. (2026). Contact dynamics characteristics of a deepwater drilling riser-drillstring coupled system. Ocean Engineering, 366, 127458.

Xu, Yuanchao, et al. "Reskoopnet: Learning koopman representations for complex dynamics with spectral residuals." arXiv preprint arXiv:2501.00701 (2025).