ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-4/W8-2025
The 8th ISPRS Geospatial Conference 2025, 15–17 December 2025, Tehran, Iran

Quantitative Uncertainty Analysis of Monocular Point Clouds for Twinning Roads

Shirin Malihi[1*], Ioannis Brilakis [1], Francesco Nex [2], Mahmoud Reza Delavar [3], Abbas Kiani [4], George Economides [5], Graham Starkey [6], Stefano Cavazzi [7], Simon Cumming [8]

[1] Civil Engineering Department, University of Cambridge, United Kingdom – sm2852@cam.ac.uk, ib340@cam.ac.uk

[2] Faculty of Geo-Information Science and Earth Observation (ITC), Netherlands - f.nex@utwente.nl

[3] Advisor of the Working Group IV/2 ISPRS; School of Surveying and Geospatial Eng., College of Engineering, University of Tehran, Iran – m.delavar@gmail.com

[4] Babol Noshirvani University of Technology, Babol, Iran - a.kiani@nit.ac.ir

[5] Digital Twins Department, UK Government's Department for Transport, United Kingdom - george.economides@dft.gov.uk

[6] BIM Department, Costain, United Kingdom - graham.starkey@costain.com

[7] Innovation and Research Department, Ordnance Survey, United Kingdom - stefano.cavazzi@os.uk

[8] Data Science Department, Capco, United Kingdom - simmon.cumming@capco.com

* Corresponding author

KEY WORDS: Gaussian distribution, Monocular point cloud, Confidence level, Road, Depth estimation

ABSTRACT

Digital twinning of road infrastructure enables the simulation of road asset conditions and performance, supporting more effective planning, management, and maintenance of road networks. RGB cameras offer a low-cost sensing solution for digital twinning applications. Monocular RGB images, in particular, provide a cost-effective source of data for road surface analysis and inspection. From these images, point clouds can be generated to reconstruct road surfaces and surrounding features.

A key factor in the reliability of digital twins is the confidence in the data sources used, especially the point clouds derived from monocular images. This confidence is closely linked to the uncertainty inherent in such reconstructions. In this paper, we analyse the uncertainty of image-derived point clouds generated using two state-of-the-art depth estimation models: Metric 3D V2 and Zoedepth. Unlike previous studies that focus mainly on geometric accuracy or semantic segmentation, our approach explicitly quantifies the statistical uncertainty and confidence intervals of monocular reconstructions at the object level, thereby introducing a systematic method for reliability assessment in road digital twinning.

We evaluate these point clouds across five distinct object clusters: pavement, tree trunks, tree crowns, lamp posts, and curbs. The analysis leverages a set of geometric features to characterize and differentiate the clusters. Confidence levels for the extracted clusters are quantified using statistical techniques based on normal distribution modelling, allowing us to assess the reliability of classification outcomes.

Our findings indicate that the confidence levels of the identified clusters, extracted from monocular point clouds using a combination of geometric descriptors and supervised machine learning techniques range between 66% and 91%. These results demonstrate that, despite the inherent limitations of monocular depth estimation, meaningful object-level segmentation is achievable with reasonable certainty. Furthermore, the variability in confidence levels highlights the differing levels of geometric distinctiveness and structural complexity among object types.

These results not only demonstrate the practical viability of monocular depth estimation for road asset modelling but also establish a quantitative baseline for uncertainty propagation in data-driven digital twins.

1. Introduction

Digital twin of roads as a digital replica of road assets should be trustable for users in different areas of traffic management, maintenance, construction and transportation. Considering the facts that 246,700 miles of motorways and roads in the UK carry over 83% of all passenger miles travelled and are responsible for 69% of all transport Greenhouse Gas Emissions (GHG) of UK, the importance of roads is clear as a critical component of infrastructure, connectivity and daily life (Malihi et al., 2024). Facilitating trade and commerce, supporting businesses, linking urban and rural areas, public transport, and emergency services are a number of domains of road significance. Digital twinning of roads can enable smarter urban planning, predictive maintenance, asset management, noise analysis, and emergency management.

Digital twin is based on the geometric model of the environment, in this case, roads. 3D modelling of road networks employs different data modalities that provide a detailed holistic view of road assets (Shokri et al., 2025). RGB images are a key modality in road modelling, providing visual information and colour, detailed representation of road conditions and edges, and road surface classification. However, RGB images are subject to variability in brightness, e.g. shadows, colour, so the accuracy of the derived point cloud can be reduced.

3D reconstruction of surfaces uses SFM (Structure From Motion) technique for a series of overlapping images to output point cloud and texture (Iheaturu et al., 2022). This photogrammetric method detects common features across multiple images using mostly SIFT (Scale-Invariant Feature Transform) or SURF (Speeded-Up Robust Features) and creates camera position and orientation for each image besides a dense point cloud and mesh of the scene. In addition to this accurate and relatively scalable method, in computer vision and robotics learning-based methods have been employed for depth estimation. Learning-based depth estimation methods leverage data-driven approaches to recover 3D geometry in many areas, including autonomous driving, 3D reconstruction, and robotics. Monocular depth estimation is a low-cost method to generate

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-479-2026

3D information from 2D data. However, depth estimation from single images in the short range is an ill-posed problem due to depth ambiguity, unknown intrinsic and extrinsic camera parameters, colour and illumination variability. On the other hand, deep learning-based methods of depth estimation formulate it as a learning problem and face with challenges of acquisition of 3D ground truth, adaptation to unseen images, high expenses in memory and time of computation. Usually these methods consist of masking, feature learning, feature selection, a combination of convolutional, pooling, and fully connected layers for coarse depth estimation, and a combination of convolutional, pooling, and concatenation layers for fine depth estimation and post-processing (Ganj et al., 2024) .

Metric depth has practical utility in 3D reconstruction, mapping, planning and robotics. However, training a single model can not be generalized for different datasets, so metric models suffer from scale ambiguity, and different target objects. Relative depth models capture the structure of the scene by indicating the relative distances of objects. For precise depth, relative models are fine-tuned for the special datasets. Metric3D v2 is a metric network and a geometric foundation model for zero-shot metric depth estimation. It deals with the metric ambiguity by transforming all images and depth labels into a canonical camera space to avoid confusing networks. Depth models were trained over 16 million of images from thousands of camera models, resulting in zero-shot generalization (Hu et al., 2024). The authors propose a transformation module that standardizes images into a canonical camera space. This approach enables the model to learn consistent depth representations across diverse camera models, facilitating accurate metric depth. By leveraging the geometric relationship between depth and normals, the model extracts knowledge from depth data to enhance normal predictions, even in the absence of explicit normal labels.

ZoeDepth is a hybrid model which pre-trained on 12 datasets using relative depth estimation. This stage focuses on learning the geometric structure of scenes without considering absolute scale, enhancing the model's generalization capabilities across various domains. Next, the model was fine-tuned on two domains (NYU and KITTI) by integrating the strengths of both relative and metric depth estimation methods. This stage calibrates the model to predict depth in real-world units, ensuring accurate scale in its estimations. During inference, a latent classifier automatically routes each input image to the appropriate domain-specific head (e.g., indoor or outdoor). This mechanism enables the model to adapt its predictions based on the characteristics of the input image, enhancing performance across different environments. A lightweight head is used with a bin adjustment design module for each domain. This model achieves unprecedented zero-shot generalization performance to 8 unseen datasets from both indoor and outdoor domains (Farooq et al., 2023). These models' ability to produce accurate metric depth maps from single images has significant implications for enhancing scale accuracy in SLAM systems by providing reliable depth estimates, mitigating scale drift issues.

In recent years, the amount of digital data collected through various types of sensors have reached a remarkable size and are growing exponentially. The acquisition properties of the data has to be extended with quality indicator characteristics for the selected domain. The ground truth is frequently no longer available, and the assessment of data quality often results in the assessment of the data source. Data quality is classified in terms of intrinsic, contextual, representational, and accessibility dimensions (Wangand Strong 1996), and affects decision quality. Intrinsic data quality is an essential component in quality assurance model and data management standards such as ISO 8000 for reliable decision-making. Uncertainty declares the doubt of measurements (Rios, 2020). Usually uncertainty has a close connection with the confidence level . Point clouds cover a surface with a certain level of uncertainty too. Figure 1 presents a diagrammatic overview of recent research on uncertainty modelling and the digital twinning process, the data, challenges and key achievements. This brief review encompasses studies published between April 2023 and May 2025, focusing on key topics such as laserscanning, building detection, Photogrammetry and Mobile Mapping Systems , semantic segmentation, Defect detection, road planning and NeRF (Hartmann, J., and Alkhatib, H., 2023, Li, J., et al. 2025, Lee, E., et al. 2024, Landgraf, S., et al. 2024, Yan, y., et al. 2025, Jäger, M., et al. 2025, Buuveibaatar, M., et al.2025, Zhixin, C., et al. 2025, Maculotti, G., et al. 2025, Haverkamp, P., et al. 2025).

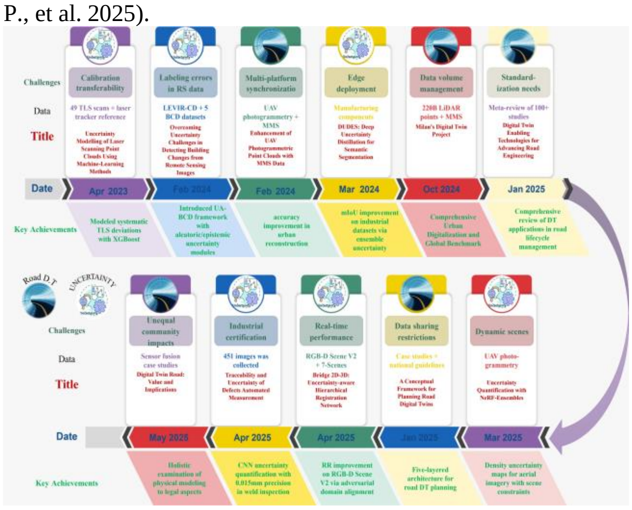


Figure 1. A diagrammatic overview of recent research on uncertainty modelling and the digital twinning process

In this research, the uncertainty of monocular depth estimation models is analysed for covering road pavement which is the base of road digital twin. Uncertainty in civil-related subjects suffers from a lack of concentrated research and application domain metrics. Usually research on uncertainty deviates to other data quality dimensions. The significance of uncertainty analysis emphasizes the trustability of decisions based on the captured data (Lark et al., 2022, Malihi 2025). Even data with high qualities of accuracy and precision is not necessarily trustable.

Considering the importance of the digital replica of roads and its geometric base, uncertainty analysis of point clouds as a sensing method is critical for road pavement. This paper consists of methodology section in which details of the devised approach including depth estimation, confidence estimation and clustering are represented; Results section in which clustering outputs and statistical analyses are enumerated; Discussion section in which uncertainty is examined and conclusions in which main consequences and future directions are discussed.

2. methodology

Monocular images of local roads around the Civil Engineering Department of the University were captured using an RGB camera mounted on the Pointpix platform (Trzeciak and Brilakis, 2022), with a speed lower than 20 km/h.
The data collection followed a structured protocol to ensure consistent lighting and geometry conditions, minimizing illumination bias and perspective distortion that typically affect monocular depth estimation accuracy.
Figure 2 illustrates the development of two deep learning-based depth estimation methods. It highlights key data, milestones, and advancements made between February 2023 and May 2025. It focuses on the initial method of ZoeDepth, addressing scale ambiguity, enhancing privacy protection, and integrating geometric priors for accurate metric depth estimation. The second method, Metric3D-V and its subsequent version, are highlighted besides integration with downstream applications, optimization for computational efficiency, and achieving a balance between accuracy and performance in depth estimation.

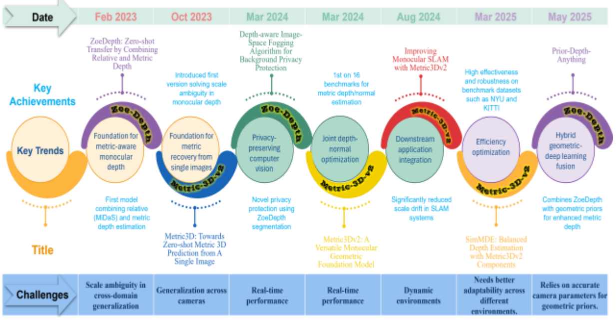
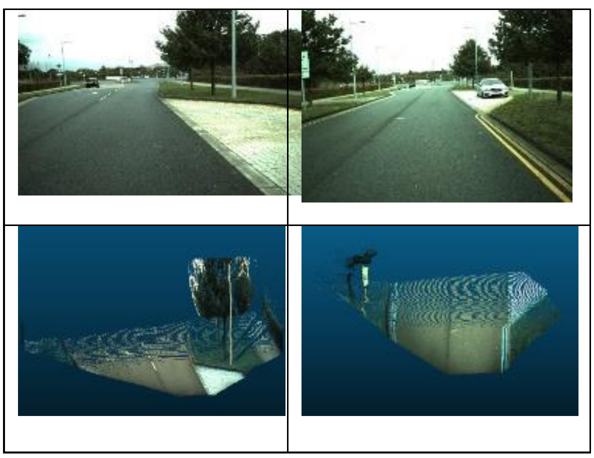


Figure 2. Progress of Zoedepth and Metric 3D

These images were tested by Zoedepth and Metric 3D v2 models, afterward, calibration reports of the camera direct to point clouds of road scenes. The focus of this research is on the trustability of this kind of image-derived point cloud mostly for pavement coverage. The selection of two fundamentally different monocular depth architectures (metric versus hybrid relative-metric) enables a comparative evaluation of uncertainty behaviour across distinct learning paradigms, thus ensuring generality in the derived findings.
In Table 1, several samples of images and resultant point clouds from Metric 3D V2 and Zoedepth are depicted. Considering the maximum and minimum focus distances of the camera, these clouds were trimmed:



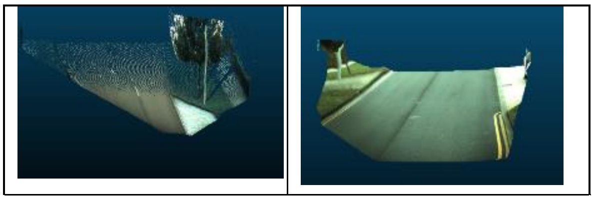


Table 1. RGB images and generated point cloud from depth estimation by Metric 3D V2

We employ geometric features for the classification of tree trunk and crown, lamp post, pavement and curb. These categories were chosen deliberately because they represent a wide geometric spectrum — planar, linear, volumetric, and irregular structures — providing a comprehensive basis for assessing uncertainty sensitivity across object geometries.

The pipeline starts with the computation of local neighbourhoods for each point and geometric features are extracted considering the spatial characteristics around the 3D point. These features are used as input for an unsupervised learning classifier to extract trained classes. We employ a filter-based feature selection method to extract linearity (*Li*), planarity (*Pi*), verticality (*Vi*), and surface variation (*Svi*).

We investigate data-driven criteria to analyse the uncertainty of extracted classes from monocular point clouds. Machine learning methods can categorize objects using a scoring approach. Considering the prediction scores confidence level of classified points is determined.

We assume feature values follow the normal distribution, as it represents a vast range of phenomena. Although the assumption of normality is a simplification, it serves as a consistent statistical framework to express confidence levels in the absence of complete ground truth, aligning with established uncertainty modelling practice in photogrammetry and geospatial data quality analysis.

Confidence levels of the 3-sigma rule express probabilities of distributed normal estimation that lie within an interval estimate. 68%, 95%, and 99% of estimated scores are expected to lie within (μ- 1σ), (μ- 2σ), (μ- 3σ), respectively, in where μ and σ are the mean and standard deviation.

In the next section results of clustering and statistical analyses of this approach are discussed.

3. Results

After pre-processing of monocular point clouds of 2 networks and noise filtering, the clouds are structured for spatial analyses. We use KDTree on account of its speed and performance capabilities for large datasets. This structuring method can organize the Euclidean distance of its nodes efficiently. This distance in the feature domain is categorized for unsupervised clustering. Some samples of extracted clusters for both depth estimation methods are displayed in Table 2.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-479-2026

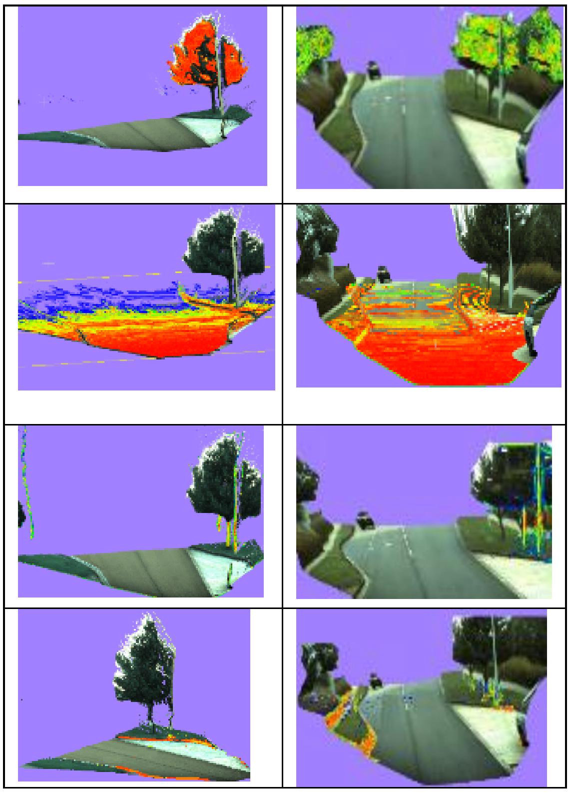


Table 2. colorized classes based on prediction scores, 1st row outputs of Metric 3D v2 , 2nd row Zoedepth

In each cluster standard deviation is examined, which is listed in Tables 3. For each cluster, a feature that gains the best results is used. Extracted point clouds are narrow view, hence the number of clusters in each cloud is limited to the intended clusters, mostly.

class		σl	σp	σv	σsv
tree trunk	Metric 3D V2	0.08			
	Zoedepth	0.09			
tree crown	Metric 3D V2		0.058	0.15	
	Zoedepth		0.068	0.2	
Lamp post	Metric 3D V2			0.045	
	Zoedepth			0.041	
pavement	Metric 3D V2		0.05		
	Zoedepth				0.01
curb	Metric 3D V2	0.1			
	Zoedepth	0.15			

Table 3. Computed standard deviation of geometric features for each class of monocular point cloud of Metric 3D V2 and Zoedepth

Distribution functions of extracted clusters are analysed and displayed in Fig. 3-11 for both methods. The mean parameter is determined using the flowcharts.

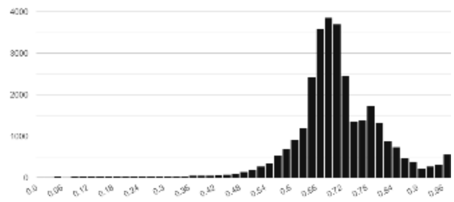


Figure 3. Metric 3D v2, trunk

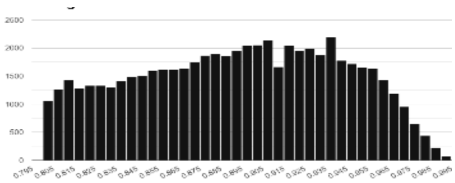


Figure 4. Metric3D /V2 crown

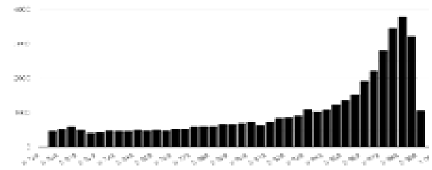


Figure 5. Metric 3D V2, Pavement

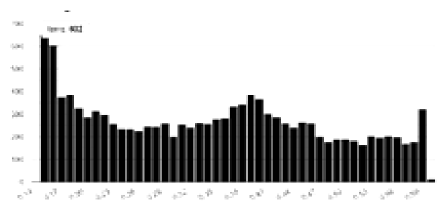


Figure 6. Metric 3D V2, Curb

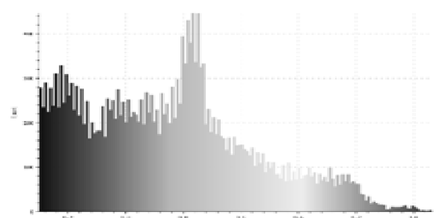


Figure 7. Zoedepth, Curb

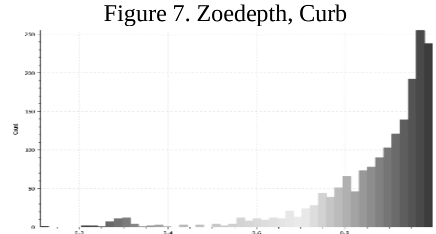


Figure 8. Zoedepth, Pavement

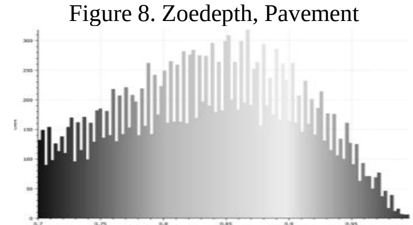


Figure 9. Zoedepth, crown

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-479-2026 | © Author(s) 2026. CC BY 4.0 License.

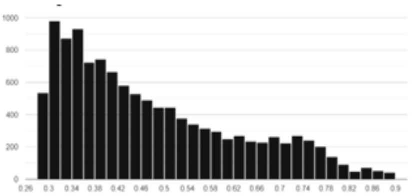


Figure 10. Zoedepth, trunk

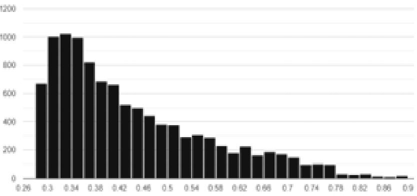


Figure 11. Zoedepth, Lamp

In the next section, details of probabilistic measurements will be discussed.

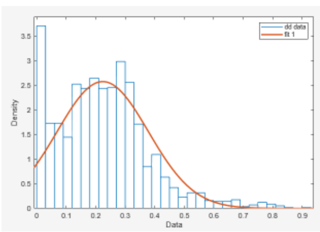
4. Discussions

For each flowchart, the normal distribution function is computed using the equation (1):

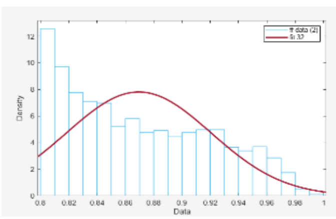
$$fx = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} e^{-\frac{(x-\mu)2}{2\sigma^2}} \qquad (1)$$

To evaluate the confidence of the results based on normal distribution, number of points outside the *fx* is computed for each flowchart. For Figures 2-10, distribution functions are examined in the following figures.

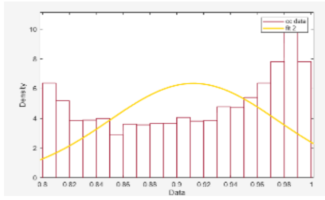
Considering the normal distribution, confidence interval and statistical parameters of the computed features before changing the scale (μ, σ), plus the population of each cluster (n), the trustability of the computed features is computed in each diagram (Figure. 12).



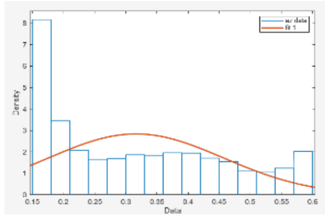
In Figure 2, 0.14 of data



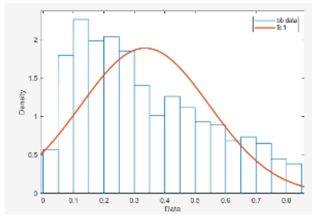
In Figure 3, 0.17 of data



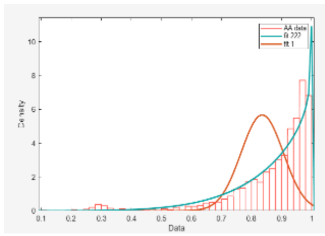
In Figure 4, 0.21 of data



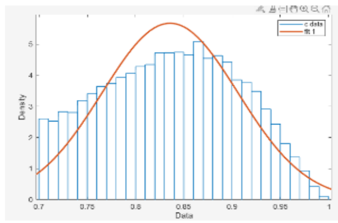
In Figure 5, 0.20 of data



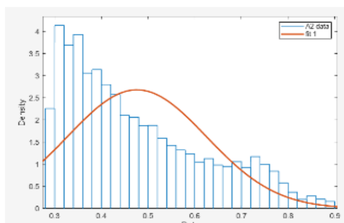
In Figure 6, 0.19 of data



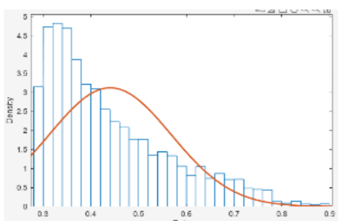
In Figure 7, 0.34 of data



In Figure 8, 0.09 of data



In Figure 9, 0.20 of data



In Figure 10, 0.15 of data

Figure 12. Distribution functions for different clusters

According to the examined ratios for each dataset, extracted clusters using the geometric features can achieve confidence levels among 66 – 91 %. This quantitative range is consistent with the theoretical expectations of uncertainty propagation in monocular depth estimation and confirms that even with a single image input, a statistically meaningful degree of trustability can be achieved.
One of the important points that affects this matter negatively is the normal distribution assumption. However, the observed deviations also provide valuable diagnostic insight into where monocular reconstruction fails — specifically for thin or reflective objects — and can inform future weighting strategies in digital twin confidence layers.
Obviously for some diagrams normal distribution is not the best choice. For example, Fig. 7 follows the Beta distribution. Furthermore, the distribution of some of the clusters consist of multiple normal functions, such as Fig. 4. Extracted uncertainty is a result of different methods of the pipeline including filtering, structuring, features, and clustering; along with uncertainty of point clouds which are generated knowledge from the depth estimation algorithm and calibration process. A comprehensive understanding of the point clouds' uncertainties is needed to investigate the uncertainties of the whole path of the approach.

5. Conclusions

In this research, a new perspective on data quality was presented. The confidence of the output from the viewpoint of geometrical analysis of point clouds of pavement was analysed.
The contribution of this study extends beyond object classification: it establishes a reproducible statistical protocol for uncertainty quantification in monocular point clouds, bridging the gap between computer vision outputs and civil engineering reliability standards.
We investigate the reliability of clustered points for various road asset classes, namely pavement, tree trunk, lamp post, tree crown, and curb, under the assumption of a normal distribution.
By linking geometric feature variability with probabilistic confidence, the study provides the first empirical baseline for evaluating uncertainty in low-cost, image-based road digital twins. Our findings suggest that monocular point clouds provide a narrow-view, 2.5-dimensional representation of pavement that adheres to the 3-sigma rule, indicating a statistically reliable coverage for the selected asset types. Our findings indicate that the confidence levels of the identified clusters, extracted from monocular point clouds range between 66% and 91%. This measurable confidence range represents a crucial step toward integrating uncertainty metadata directly into digital twin pipelines, allowing future users to make informed, risk-aware decisions when relying on monocular datasets.
This evaluation framework serves as a foundation for refining point cloud segmentation pipelines and can inform further development in urban scene understanding, infrastructure monitoring, and autonomous navigation systems.

To enable practical, real-world applications, techniques such as data compression, dimensionality reduction, and data fusion are recommended. Future works will explore alternative statistical distributions and probabilistic strategies to more accurately evaluate the trustworthiness of the derived data. Furthermore, a dedicated ground truth is essential for assessing uncertainty at various stages of the processing pipeline. To support this, new datasets will be developed to benchmark the uncertainty chain comprehensively.

Acknowledgement

This research is supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie grant agreement No. 101034337.

References

Farooq, S., Kaust, B., Birkl, R., Wofk, D., Wonka, I. P., & Müller, M., 2023. ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth. https://arxiv.org/abs/2302.12288v1

Hu, M., Yin, W., Zhang, C., Cai, Z., Long, X., Wang, K., Chen, H., Yu, G., Shen, C., & Shen, S., 2024. Metric3Dv2: A Versatile Monocular Geometric Foundation Model for Zero-shot Metric Depth and Surface Normal Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2024.3444912

Malihi, S., Potseluyko, L., Mathew, A., Alavi, H., Reja, V. K., Pan, Y., Binni, L., Guangming Wang, Xiang Wang, & Brilakis, I., 2024. Review of multimodal data and their applications for road maintenance. Smart Construction. https://doi.org/10.55092/SC20240010

Iheaturu, C., Okolie, C., Ayodele, E., Egogo-Stanley, A., Musa, S., Speranza, C., A., 2022. simplified structure-from-motion photogrammetry approach for urban development analysis, Remote Sensing Applications: Society and Environment, https://doi.org/10.1016/j.rsase.2022.100850

Lark, R.M., Chagumaira, C., Milne, A.E., 2022. Decisions, uncertainty and spatial information, Spatial Statistics, Volume 50, https://doi.org/10.1016/j.spasta.2022.100619.

Trzeciak, M., Brilakis, I., 2022. Precise and dense AI-based mobile 3D reconstruction of indoor scenes by camera-lidar fusion and odometry, Proceedings of the 2022 European Conference on Computing in Construction.

Malihi, S., 2025. Multimodal data, Confidence and Digital Twin, World Journal of AI and Robotics

Rios, J., Staudter, G., Wber, M., Anderl, R., Bernard, A., 2020. Uncertainty of data and the digital twin: a review, International Journal of Product Lifecycle Management

Haverkamp, P., Anantheswar, A., May, M., Walding, G., Hildebrandt, J., Thiessat, K.A., Wollny, I., Traverso, M., Müller-Mall, S., Kaliske, M., Habich, D., Lehner, W., 2025. Digital Twin Road: value and implications involving data and application, Discover Applied Sciences.

Maculotti , G., Giorio, L., Genta, G., Galetto, M. 2025. Traceability and uncertainty of defects automated measurements by CNN-powered Machine Vision Systems, CIRP Annals

Cheng, Z., Deng, J., Li, X., Yin, B., Zhang, T., 2025. Bridge 2D-3D: Uncertainty-aware Hierarchical Registration Network with Domain Alignment, Arxiv

Buuveibaatar, M., Brilakis, I., Peck, M., Economides, G., Lee, W., 2025. A Conceptual Framework for Planning Road Digital Twins, Buildings

Jäger, M., Landgraf, S., Jutzi, B., 2025. Density uncertainty quantification with NeRF-Ensembles: Impact of data and scene constraints, International Journal of Applied Earth Observation and Geoinformation

Yan, Y., Ni, L., Sun, L., Wang, Y., Zhou, J., 2025. Digital Twin Enabling Technologies for Advancing Road Engineering and Lifecycle Applications, Engineering

Landgraf, S., Wursthorn,, K., Hillemann,, M., Ulrich,, M., 2024. DUDES: Deep Uncertainty Distillation using Ensembles for Semantic Segmentation, PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science

Lee, E., Park, S., Jang, H., Choi, W., Sohn, H. 2024. Enhancement of low-cost UAV-based photogrammetric point cloud using MMS point cloud and oblique images for 3D urban reconstruction, Measurement

Li, J., He, W., Li, Z., Guo, Y., Zhang, H., 2025. Overcoming the uncertainty challenges in detecting building changes from remote sensing images, ISPRS Journal of Photogrammetry and Remote Sensing

Hartmann, J., Alkhatib, H., 2023. Uncertainty Modelling of Laser Scanning Point Clouds Using Machine-Learning Methods, Remote Sensing

Rios, J., Staudter, G., Weber, M., Andrel, R., Bernard, A., 2020. Uncertainty of data and the digital twin: a review, International Journal of Product Lifecycle Management

Shokri, D., Malihi, S., Homayouni, S., Larouche, C., Rastiveis, H., Brilakis, I., Pole-Shaped Object Extraction from LiDAR Point Clouds of Roads, 2025 Proceedings of the 42nd ISARC, Montreal, Canada.