

Railway track surface defect 3D reconstruction and relocalisation through cross-modality fusion

Yizhong Wang , Lichao Yang, Isidro D. Cardenas, Yifan Zhao *

Faculty of Engineering and Applied Sciences, Cranfield University, Bedford MK43 0AL, United Kingdom

ARTICLE INFO

Keywords:

Augmented reality
Multi-sensor Information Fusion
3D registration
Rail defect relocalisation

ABSTRACT

Accurate localisation of track surface defects is critical for safe and cost-efficient railway maintenance. However, current reinspection workflows using portable non-destructive testing devices face positioning uncertainty due to the absence of spatial anchors. This research proposes a relocalisation workflow that requires neither ground infrastructure nor depth sensors. In the inspection stage, a three-layer 3D multimodal track model is reconstructed from digital images and thermograms using structure-from-motion and 2D-2D cross-modality registration. During the relocalisation stage, perspective-n-point is employed for 3D-2D spatial registration, projecting the 3D model onto on-site images to achieve accurate defect localisation without reinspection. The method was evaluated using nine shooting positions on seven artificial defects. Projected defects demonstrated improved perceptual separability (edge strength and contrast-to-noise ratio) over on-site images. At optimal imaging positions, the relocated defect centroids exhibited displacement errors range from 0.34 mm to 3.62 mm. The projection accuracy was further evaluated by Intersection over Union values ranging from 0.961 to 0.993 and Structural Similarity Index Measure ranging from 0.845 to 0.968.

1. Introduction

Railway track maintenance incurs significant expenditure due to its direct impact on safety, service continuity, and infrastructure reliability. In the 2023–2024 period, Network Rail spent £32.3 billion on infrastructure maintenance and renewal, accounting for 43% of its total expenditure (Office of Rail and Road, 2024). Despite this extensive spending, track defects, including broken rails and track-geometry irregularities, are the dominant contributors to derailments, representing approximately 20% of all cases (Liu et al., 2011; Zhang et al., 2022). Also, maintenance activities result in up to 1,400 h of train delays per week, accounting for approximately 70% of total railway delays (Winsor et al., 2000). With the emergence of digital maintenance paradigms (Scheffer et al., 2021; Mikusova et al., 2024a), established maintenance workflows are increasingly being integrated with information technologies towards intelligent, digitalised, and visualised processes, with the aim of reducing costs and improving efficiency. As shown in Fig. 1, the current operational technology (OT)-centric workflow (Keylin, 2019; Kooban et al., 2023) comprises four stages: (1) acquisition of non-destructive testing (NDT) signals along with corresponding Global Positioning System (GPS) coordinates; (2) defect diagnosis based on

NDT signal analysis; (3) coarse GPS-based navigation; and (4) precise on-site defect relocalisation using portable NDT devices. Within this workflow, existing digitalisation research mainly focuses on 3D reconstruction of railway tracks, multimodal sensor information fusion (MSIF) of NDT data, and relocalisation of track defects.

1.1. 3D reconstruction

In the railway domain, mainstream non-contact 3D reconstruction approaches for railway tracks can be categorised into passive and active methods. Passive methods recover implicit geometric structures from camera observations through algorithmic optimisation, such as structure-from-motion (SfM) (Gabara and Sawicki, 2018). In terms of camera configurations, multi-camera-based approaches have achieved sub-millimeter reconstruction accuracy (Zhang et al., 2023a,b), compared with the centimetre-level accuracy typically obtained by monocular methods (Mahmoud et al., 2022). Further experimental studies demonstrated that camera placement, shooting intervals, and working distances, directly affects reconstruction accuracy, with a best-case reconstruction error of 0.4 mm reported (Wang, et al., 2024a,b). By contrast, active methods generate depth maps by actively emitting signals in specific spectral bands, such as laser-based 3D scanning (Chen

* Corresponding author.

E-mail address: yifanzhao@ieee.org (Y. Zhao).

<https://doi.org/10.1016/j.isprsjprs.2026.03.021>

Received 14 September 2025; Received in revised form 5 February 2026; Accepted 14 March 2026

Available online 25 March 2026

0924-2716/© 2026 The Author(s). Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature		IR	Infrared
1D	One-Dimensional	MSIF	Multimodality Sensor Information Fusion
2D	Two-Dimensional	NDT	Non-Destructive Testing
3D	Three-Dimensional	OT	Operational Technology
AR	Augmented Reality	PLA	Polylactic Acid/ Polylactide
BA	Bundle Adjustment	PMVS	Patch-based Multi-View Stereo
CNC	Computer Numerical Control	PnP	Perspective-n-Point
CNR	Contrast-to-Noise Ratio	PSNR	Peak Signal-to-Noise Ratio
CVMS	Clustering Views for Multi-view Stereo	RGB	Red, Green, Blue
DLT	Direct Linear Transformation	RTK	Real-Time Kinematic
DPC	Dense Point Cloud	SfM	Structure-from-Motion
ES	Edge Strength	SIFT	Scale-Invariant Feature Transform
FOV	Field of View	SPC	Sparse Point Cloud
GNSS	Global Navigation Satellite System	SSIM	Structural Similarity Index Measure
GPS	Global Positioning System	SVD	Singular Value Decomposition
IMU	Inertial Measurement Unit	UT	Ultrasonic Testing
IoU	Intersection-over-Union	UWB	Ultra-Wideband

et al., 2022; Liu, et al., 2024a). Laser 3D scanning offers superior modelling accuracy (as high as 0.02 mm (Liu, et al., 2024a)) and faster reconstruction speed (Bianco et al., 2013) compared with SfM-based methods. These advantages stem from fundamentally different reconstruction principles: SfM-based approaches rely on feature correspondences extracted from images with rich surface texture, whereas laser-based systems directly measure distances using dedicated sensors, with cameras primarily used for colour mapping. Therefore, images used in the SfM-based methods inherently provide traceable texture information for localisation, while point clouds generated by laser-based methods lack such visual texture details. Additionally, laser scanning systems typically involve bulky hardware and complex calibration processes due to the integration of additional laser sensors. From the cost perspective, a typical commercial laser scanning system costs approximately \$100,000, while a camera-based system costs only around 6–8% of this amount (Xue et al., 2021).

1.2. Multimodality data fusion

NDT technologies have been widely applied in railway inspection, as comprehensively reviewed in recent studies (Xiong et al., 2023; Liu et al., 2025; Mordia and Verma, 2025). To enhance the reliability of defect detection and diagnosis, multiple NDT modalities have been integrated into unified information representations, enabling cross-validation of defect indications (Miles et al., 2024; Zhang et al., 2024). Among these approaches, the fusion of camera-based visual inspection and thermography is particularly common due to their similar imaging principles and 2D image representations. Existing studies combining visual inspection and passive thermography are mainly used in large-scale scenes or scenarios with naturally occurring thermal sources, such as obstacle detection (Xu et al., 2023; Yang et al., 2024), cable terminal inspection (Liu, et al., 2024b) and vehicle bearing inspection (Liu et al., 2022). Unlike passive thermography, active thermography

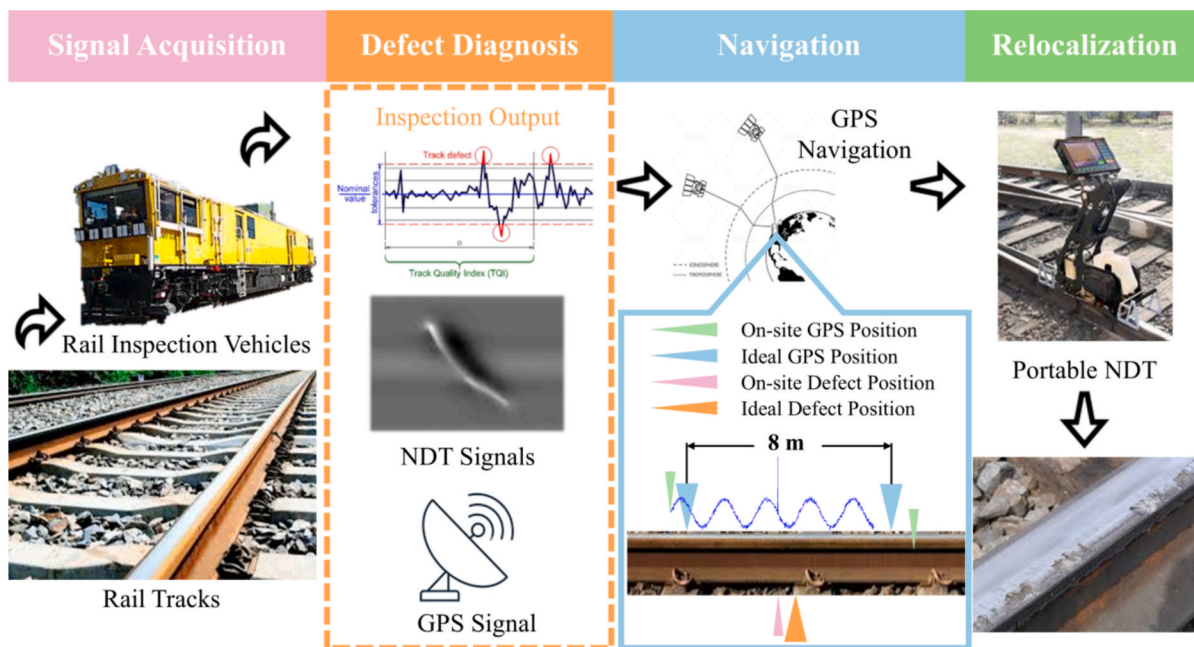


Fig. 1. The four-stage workflow of rail track defect inspection and relocalisation. The blue box represents the defect relocalisation stage based on GPS error with pre-inspection NDT signals along with the longitudinal direction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

generates a pronounced temperature contrast between defective and non-defective regions by heating object surfaces. Owing to the intuitive and human-interpretable nature of both thermal and visible imagery, image-level fusion studies predominantly focus on improving detection accuracy, such as railway ballast fouling detection (Koozhmishi and Connolly, 2025), train braking component detection (Giulietti et al., 2025) and rolling contact fatigue detection (Wang et al., 2026). At the 3D level, a GRIF (geometric, RGB, infrared, and fusion) model was recently proposed as a multimodal information representation for railway tracks with a reprojection error of 0.4 mm (Wang et al., 2025). However, its spatial accuracy and resolution are generally inferior to those of the corresponding 2D images due to reconstruction and projection errors.

Despite these advancements, current research focuses on individual stages of the maintenance workflow in isolation, without explicitly modelling or exploiting the data connections and information flow across the four stages. Specifically, the high-performing NDT visualisation results are only used to assess health conditions of railway tracks and cannot assist in on-site damage relocalisation and visualisation. The primary reason is the lack of spatial anchoring caused by misalignment between the relative coordinates of railway inspection vehicle signals and the absolute physical coordinates of railway track. Because of the limited positioning resolution of GPS (typically in the range of 6–10 m (Zhou et al., 2025), and up to 20 m in urban areas (Wu et al., 2021)), relocalisation of damaged areas mainly relies on reinspection using portable NDT equipment, such as SonicVizio-WS-S. This disconnection, when decomposed along the inspection workflow, originates from two distinct sources: the macroscopic GPS-based navigation process and the microscopic defect relocalisation process. GPS-based navigation only provides a coarse spatial window within which defects exist, with positioning errors present in all three directions: longitudinal, transverse, and vertical. For example, as shown in Fig. 1 (the blue box), the longitudinal positioning is affected by errors. With GPS errors only considered, the uncertainty of this window increases, making it indeterminate on which side the actual defect position lies relative to the estimated GPS coordinate. At the microscopic signal scale, although NDT techniques provide sub-millimeter diagnostic accuracy (Pham et al., 2021; Xue et al., 2023; Mussatayev and Kempka, 2024), using GPS as the sole spatial reference causes the defect locations indicated in the NDT signals to deviate from their physical positions along the track, as illustrated by the blue box in Fig. 1.

1.3. Damage relocalisation

To address the location uncertainty caused by GPS, the mainstream approach is to improve GPS resolution by deploying fixed and vehicle-mounted instruments. Recent research has focused on improving the accuracy and robustness of GPS-based positioning in railway applications by applying additional spatial anchors. Mikhaylov et al. (2023) proposed a Global Navigation Satellite System (GNSS) framework augmented with an Inertial Measurement Unit (IMU) and Real-Time Kinematic (RTK) reference stations, achieving an average positioning error of 1.3 m in dynamic conditions. Wang et al. (2024a) developed a real-time train positioning system by integrating Ultra-Wideband (UWB) anchors and IMU via an Error-State Kalman Filter, reaching an average error of 0.425 m in urban transit environments at 36 km/h. While these studies demonstrated promising positioning performance, their reliance on ground infrastructure, such as RTK-based stations and UWB tunnel anchors, increases deployment complexity and cost in large-scale railway scenarios. Notably, this spatial resolution limitation cannot be fully addressed by simply increasing the GPS sampling frequency, as the dominant errors come from systemic factors such as satellite ephemeris inaccuracies, signal propagation delays, and multipath distortions (European GSA, 2019; Mikhaylov et al., 2023). Although these techniques partially compensate for the limited resolution of GPS and reduce positioning uncertainty, they do not fundamentally eliminate the need

for repeated on-site scanning and remain insufficient to meet the millimeter-level localisation requirements. As a result, on-site engineers are still required to operate portable NDT equipment and interpret associated signals to locate and assess damage. Prolonged field operation can lead to cognitive fatigue among engineers, thereby increasing the risk of misdiagnosis during on-site inspections. Therefore, establishing a spatial reference that anchors pre-inspection signals to the physical rail target on-site is a critical step to replace the manual reinspection stage and provide accurate real-time visualisation of detected damage.

As another promising method, augmented reality (AR) overlays virtual information onto real-world assets in real-time through tracking, registration and visualisation algorithms (Bakir et al., 2023; Dargan et al., 2023; Xu et al., 2023). Al-Sabbag et al. (2022) proposed an AR system that measures the dimensional information of defects at building corners in real time using a feature back-propagating refinement scheme, with results visualised via a headset. This solution can locate defects within 2–3 s when stable and well-defined surface features are available, while the positioning error in real-time operation is close to 10%. Researchers at the University of Maryland proposed an AR tool that visualises underground pipelines with associated digital information (e.g., directions, types and material) on a smartphone (Cui and Liang, 2023). However, the virtual models displayed on the smartphone suffer from vertical drift and unstable pose updates. Additionally, AR technology has been widely explored in fields such as virtual city (Portalés et al., 2010; Hou et al., 2024), indoor localisation (Li et al., 2023; Ou et al., 2025), and medicine (Bona et al., 2022; Bruno et al., 2022), while its use in railway track inspection remains limited. Several studies explored the use of AR for defect relocalisation and visualisation for railway tracks, but lack high-resolution rendering, robust registration, or quantitative validation, limiting their reproducibility and application credibility (Porsch and Lehmann, 2023; Ahmad et al., 2024). A notable commercial example is the QoniQ Rail Explorer developed by VOSSLOH, which implements an AR system to visualise ultrasound (UT) results by mapping signal amplitude to a longitudinal rail coordinate system (VOSSLOH, 2025). Although this system provides spatial alignment between UT signals and their physical acquisition locations, the rendered outputs lack detailed surface texture and intuitive defect interpretation due to limitations in sensing modalities and visual fidelity. These limitations underscore the need for task-specific AR solutions that can support accurate defect relocalisation and informed decision-making (Rahman et al., 2023; Mikusova et al., 2024b).

1.4. Aims and contributions of this study

As a proof-of-concept, this paper proposes a multimodal reconstruction and relocalisation framework for visualising and localising railway track surface defects on-site. The framework establishes a spatial mapping process, including: (1) 2D-2D cross-modality registration between RGB and infrared (IR) images; (2) 2D-3D textured geometry reconstruction; (3) and 3D-2D projection using the estimated portable camera pose. Inspection data and on-site observations are aligned using the pre-calibrated relative pose between the RGB and IR sensors, along with the registration process that links the reconstructed 3D model to the real-world scenes. This workflow, encompassing inspection, reconstruction, and relocalisation, enables the spatial calibration and visualisation of multimodal detection results without requiring re-inspection. The proposed framework aims to facilitate accurate and efficient defect localisation, reduce on-site operation time, and improve the continuity and efficiency of data utilisation in the OT-based inspection workflow.

The main contributions of this study are as follows:

- Developed and evaluated a millimeter-level relocalisation method to accurately pinpoint railway track defects in previously unseen on-site images, mitigating GPS inaccuracies and eliminating the need for repeated defect detection.

- Proposed a cascading framework that builds registration relationships among 2D (thermal and digital) images, 3D geometric models, and on-site images through feature point matching. The framework projects geometric, RGB texture, thermogram, and fused data in 3D onto real-world scenes, enabling simplified defect assessment for railway tracks.
- Established an integrated pipeline for railway track maintenance, encompassing multimodal data acquisition and 3D reconstruction during inspection, and defect projection during on-site relocalisation, facilitating complete workflow from initial detection to in-field assessment.

2. Methodology

This section presents the proposed integration framework for surface defect reconstruction and accurate relocalisation under the GPS guidance. The framework integrates 3D reconstruction, cross-modality image registration, and relocalisation to establish spatial anchors that accurately map 3D NDT information representation into the corresponding physical parts. The framework consists of a 3D multi-sensor information fusion (MSIF) reconstruction stage and a relocalisation stage, as illustrated in Fig. 2. Two NDT technologies, including vision-based inspection and pulsed thermography, are integrated for damage assessment. Vision-based inspection enables accurate detection of surface defects while capturing rich texture information for 3D representation and spatial anchoring, supporting subsequent registration in the defect relocalisation process. Minor, shallow, or colour-similar defects are often difficult to detect in RGB images due to insufficient contrast and limitations in detection. To distinguish these defects and expand the detection range, pulsed thermography is integrated because of its strong thermal contrast after heating between healthy and defective areas. However, IR images lack background texture, making it difficult to anchor these defects in complex environments spatially. By integrating RGB and IR, the framework utilises texture and geometry in RGB images to spatially locate defects, while IR images enhance the visibility of defect regions. This combination facilitates accurate defect localisation by aligning texture-based positioning with thermography-based defect

highlighting.

During the reconstruction stage, an inspection system deployed on a vehicle operating on tracks captures RGB, active IR images, and GPS data to construct a digital 3D MSIF model, comprising layers of railway tracks that incorporate IR, RGB, and fusion data. In the relocalisation stage, guided by GPS, engineers perform coarse navigation to areas with defects for maintenance. A portable device is then employed to capture on-site images and refine the localisation of defects. Feature matching and Perspective-n-Point (PnP) estimation are performed to compute the camera pose and establish spatial anchors on-site. This process achieves accurate defect projection and relocalisation by aligning textures, thermal information, and fused layers of defects onto the actual rail surface.

Following the above rationale, the proposed solution requires a shared variable that is consistently available across all stages of the workflow and can serve to integrate the entire process. Two-dimensional RGB images constitute the fundamental data source for 3D reconstruction and MSIF, while also serving as one of the primary information representations of on-site scenes. Feature points of these images are adopted as the shared variable integrating the entire workflow. Previous studies (Tareen and Saleem, 2018) suggest that, compared with speeded up robust features (SURF), KAZE and the other three feature types, SIFT does not achieve the shortest matching time but provides the highest accuracy with respect to scale, rotation, and affine variations. Given that the proposed experiment prioritises reliable feature correspondence across viewpoint changes and long spatial baselines, rather than real-time performance, SIFT is therefore selected as the feature descriptor to ensure stable and repeatable matching throughout the workflow.

Additionally, existing studies on railway track 3D reconstruction employ SIFT-based methods, further confirming their stability and practical applicability in this domain (Zhang et al., 2023a,b; Wang, et al., 2024a,b; Wang et al., 2025). Although limited research exists on railway track registration, the parameters of reconstruction and PnP algorithms are settled based on similar research (Hesch and Roumeliotis, 2011; Sun et al., 2022). Considering the inherent characteristics of railway tracks, such as sparse feature points and high structural repetitiveness, this study adopts the PnP pipeline implemented in COLMAP

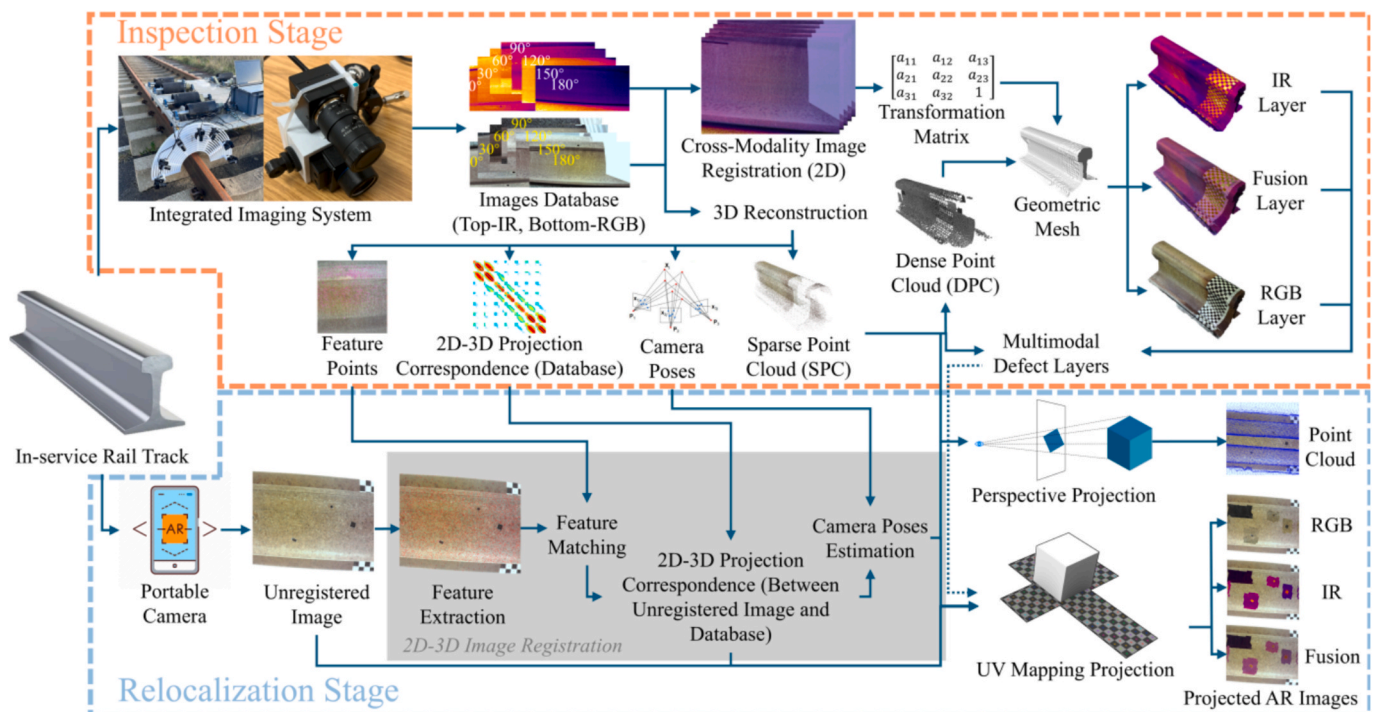


Fig. 2. The proposed 3D MSIF defect detection and relocalisation framework.

for 2D-3D registration, with specific parameters shown in Table 1.

2.1. 3D MSIF reconstruction stage

2.1.1. Experimental setup

This stage generates a sectional model by integrating RGB, IR, and fusion data into a spatially registered 3D representation of railway tracks. The reconstruction specifically targets track segments containing surface defects within the GPS-derived uncertainty zone. It would generate a visual representation providing spatial references for further defect relocalisation. As represented by the orange box in Fig. 2, an integrated imaging system is mounted on an inspection vehicle to capture RGB and active thermal images of in-service railway tracks, together with GPS coordinates. The collected multi-modal data support two tasks: (1) RGB-based 3D reconstruction via Structure-from-Motion (SfM) (Wu, 2013; Wang, et al., 2024a,b), Clustering Views for Multi-View Stereo (CMVS) (Furukawa et al., 2010; Wei et al., 2019), Patch-based Multi-View Stereo (PMVS) (Furukawa and Ponce, 2010; Yaopu et al., 2018), and Poisson surface reconstruction; (2) pixel-level cross-modality image registration to generate 3D MSIF information representation. The intermediate geometric representations, including feature points, 2D-3D projection correspondences, camera poses, and a sparse point cloud (SPC), are extracted and stored as spatial priors to support relocalisation and alignment in the next stage, as shown in Fig. 2.

Fig. 3 illustrates the experimental platform and the image acquisition process. The physical setup features a 1.4-meter CEN56E1 rail specimen, representative of the predominant rail type used in the UK, which also simulates the GPS-derived uncertainty zone. Mounted on a CNC-controlled arch, simulating the inspection vehicle, this imaging system houses a high-resolution RGB camera (3264×2248 pixels) and a FLIR A35 infrared camera (320×256 pixels, 45° field of view (FOV)), both rigidly secured using customised PLA connectors and cable ties to minimise vibration and preserve calibration accuracy during motion. To ensure consistent illumination during acquisition, two standard LED bulbs (6 W, each with 800 lm) are installed symmetrically on the arch. A high-intensity thermal excitation lamp (BW-3935, 750 Ws) is also included in the setup. It is manually triggered to emit a 1/2400 s flash pulse, which reveals subtle surface discontinuities by introducing a transient temperature differential. A checkerboard pattern composed of 10 mm grids is affixed to the rail surface to provide distinctive features for pixel-level alignment across visual and thermal modalities. Artificial

defects are introduced to assess system performance. As shown in Fig. 3 (left), these include five surface markers (defects 1–5) made from matte black paper to simulate non-structural defects, and two machined features (defects 6, 7) that emulate structural defects. To reduce the computational burden and enhance the visual prominence of the defect areas, the reconstruction focuses on regions (a length of 39 cm) surrounding the defects rather than the entire track.

As shown on the right side of Fig. 3, the integrated system captures images from seven discrete angular viewpoints (0° , 30° , 60° , 90° , 120° , 150° , and 180°) along a semi-circular arch with a 30 cm radius (shooting distance). At each angle, the system translates linearly in 1 cm increments along the rail, spanning a total distance of 39 cm to simulate the defect-surrounding area. For each shooting position, a pair of an RGB image and an IR video sequence is acquired. The second frame following the thermal flash is selected from each IR sequence to represent the thermal response, resulting in a total of 40 RGB-IR image pairs per view angle. The entire process is repeated across all seven angles, yielding a pre-inspection database of 280 registered RGB and IR image pairs for subsequent 3D MSIF reconstruction and relocalisation.

2.1.2. 3D reconstruction

In this study, SfM, CMVS/PMVS, and Poisson reconstruction are employed to perform 3D reconstruction based on RGB images with unknown camera intrinsics. There are two processes related to defect relocalisation: (1) feature point extraction and matching, and (2) triangulation.

In this feature extraction process, the Scale-Invariant Feature Transform (SIFT) (Cruz-Mota et al., 2012; Lindeberg, 2012) is utilised to extract feature points and descriptors from the captured RGB images. Each feature point searches for its nearest neighbour in the descriptor space using Euclidean distance to find the corresponding match across other images. Unstable feature point pairs are filtered out using Lowe's ratio test (Lowe, 2004; Andrianova and Demidova, 2021) with a threshold of 0.8, resulting in a set of reliable correspondences denoted as M_{RGB} in Eq. (1), written as

$$M_{RGB} = \bigcup_{(i,k)} M_{ik} = \bigcup_{(i,k)} \left\{ (p_i^j, p_k^j) | j \in J_{ik} \right\} \text{ where } p_i^j, p_k^j \in \mathbb{R}^2 \quad (1)$$

where i and k represents the indices of any two RGB images; j is the index of a matched feature point; p_i^j, p_k^j represents the 2D pixel coordinate of the point j in images i and k ; J_{ik} denotes the set of all shared point indices between images i and k . The setting of 0.8 was chosen because it has been empirically validated as optimal across diverse datasets and scenarios (Lowe, 2004).

From the 2D-2D correspondence set in M_{RGB} , the fundamental matrix is estimated via the normalised eight-point algorithm (Hartley, 1997; Chen et al., 2009) with RANSAC (Derpanis, 2010; Imre and Hilton, 2015) to represent the epipolar geometry among images. By fixing the first camera at the world origin ($P_i = [I|0]$), the epipole is extracted from the fundamental matrix to recover the projective matrix P_k and the camera pose of the second camera directly, which serves as the initialisation for subsequent 3D reconstruction processes.

The triangulation process is to calculate the 3D coordinates of feature points in the reference frame of the image i by using direct linear transformation (DLT) (Shapiro, 1978; Qin et al., 2022) and singular value decomposition (SVD). By substituting Eq. (2) into the epipolar geometry relationship, the coordinate expression of the feature point p_i^j corresponding to a 3D point X_j can be derived, as presented in Eqs. (3) and (4).

$$P_i = \begin{bmatrix} p_i^1 \\ p_i^2 \\ p_i^3 \end{bmatrix} \text{ where } \in \mathbb{R}^{3 \times 4} \quad (2)$$

Table 1

Parameters of SIFT, feature matching, RANSAC, BA and PnP algorithms.

Algorithm	Parameters	Content
SIFT feature extraction	Number of features per image	8192
	Contrast threshold	0.04
	Edge Threshold	10
	Initial Sigma	1.6
Feature Matching	Descriptor distance metric	Euclidean (L2)
	Matching strategy	Mutual nearest neighbor
	Lowe's ratio test	0.8
	Absolute distance threshold	0.7
RANSAC	Inlier threshold (F-matrix)	2
	Inlier threshold (H-matrix)	4
	Maximum iterations	1000
	Full BA step interval	Every 5 images
Bundle adjustment	Full BA max iterations	100
	Particle BA window size	20 images
	Partial BA interactions	5
	Convergence criterion	MSE < 0.25 px
PnP	Maximum Reprojection Error Threshold	25
	Minimum Number of Inliers for Absolute Pose Estimation	4
	Minimum Inlier Ratio for Absolute Pose Estimation	0.01

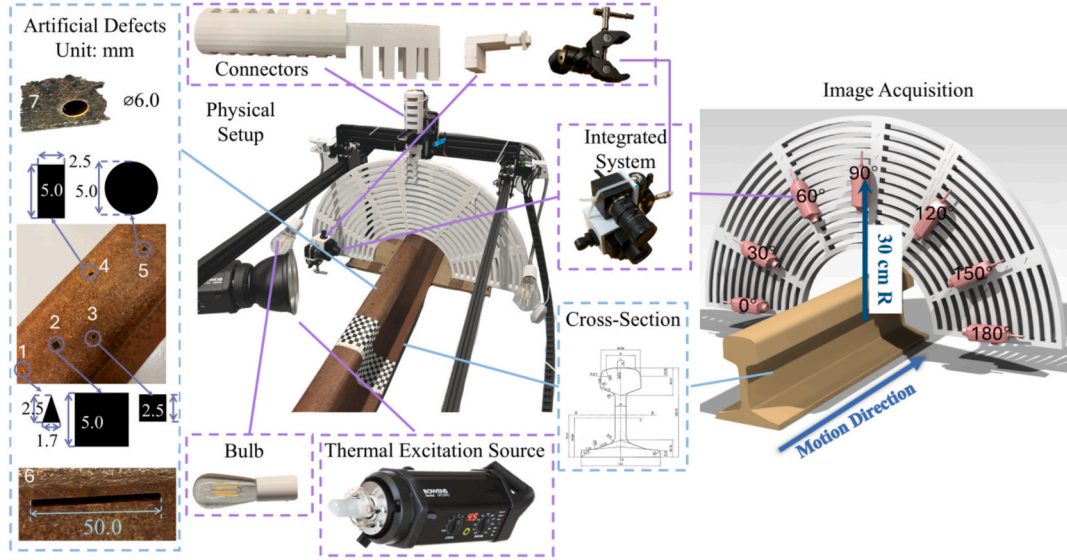


Fig. 3. An illustration of the physical experimental setup (left side) and image acquisition process (right side).

$$u_i^j = \frac{P_i^1 \bullet \tilde{X}_j}{P_i^3 \bullet \tilde{X}_j} \Rightarrow u_i^j \bullet P_i^3 - P_i^1 = 0 \quad (3)$$

$$v_i^j = \frac{P_i^2 \bullet \tilde{X}_j}{P_i^3 \bullet \tilde{X}_j} \Rightarrow v_i^j \bullet P_i^3 - P_i^2 = 0 \quad (4)$$

where (u_i^j, v_i^j) is the pixel position of p_i^j , and P_i^1, P_i^2, P_i^3 represents the first, second, and third row vectors of the projection matrix P_i . Applying the same execution to the image k yields a triangulation coefficient matrix A_j , formed using the feature correspondences between images i and k , as shown in Eq. (5). The 3D points X_j is obtained via linear triangulation using DLT and SVD. The 2D feature points p_i^j used in the triangulation process, together with corresponding 3D points X_j , form a 2D-3D correspondence $C_{2D \leftrightarrow 3D}(p_i^j, X_j)$.

$$A_j \tilde{X}_j = 0 \text{ (where } A_j = \begin{bmatrix} u_i^j \bullet P_i^3 - P_i^1 \\ v_i^j \bullet P_i^3 - P_i^2 \\ u_k^j \bullet P_k^3 - P_k^1 \\ v_k^j \bullet P_k^3 - P_k^2 \end{bmatrix} \in \mathbb{R}^{4 \times 4}) \quad (5)$$

The triangulation process is repeated with the addition of new images and feature correspondences, continuing until all available features across the image set have been reconstructed. This iterative module is known as incremental structure-from-motion (SfM). During incremental SfM, drift error accumulates with the addition of new images and feature points. To minimise this error, bundle adjustment (BA) (Engels et al., 2006; Cucci et al., 2017) is employed after a predefined number of images to refine the reconstructed sparse point cloud.

Due to the limited point density in the SPC, the clustering views for CMVS/PMVS are used to generate the dense point cloud (DPC). With its higher spatial density, this DPC also provides per-point surface normals that define the gradient field for Poisson surface reconstruction. The output of the Poisson equation is a watertight and smooth triangular mesh (Michael et al., 2006; Sellán & Jacobson, 2022) for further rendering.

2.1.3. Cross-modality image registration

In the experimental setup, the fixed relative position between the IR and RGB cameras theoretically guarantees the existence of a specific

viewpoint at which both modalities capture identical scene content, neglecting differences in FOV and lens distortion. Thus, for the RGB image captured at the shooting position s , there theoretically exists a corresponding IR image captured at position $s + \Delta s$ that shares an equivalent viewpoint and scene content. To search for the optimal offset Δs , RGB and IR image pairs were registered at varying Δs intervals manually. In practical experiments, due to the lack of distinctive and stable features on the track surface, a checkerboard pattern was affixed to facilitate registration assessment. By evaluating their corresponding Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) after binarisation, the optimal Δs is chosen as 5, as validated in our previous study (Wang et al., 2025). This process is summarised as Eq. (6). Among these functions, $R(\bullet)$ is the registration operation to generate a projected IR image, covering the transformation matrix $P_{\Delta s}$ calculation and projection achievement (Eq. (7)).

$$Q(\Delta s) = SSIM(B(ROI(I_{RGB}^s, R(I_{RGB}^s, I_{IR}^{s+\Delta s})))) + PSNR(B(ROI(I_{RGB}^s, R(I_{RGB}^s, I_{IR}^{s+\Delta s}))))s, \Delta s \in \mathbb{N} \quad (6)$$

$$I_{IR \rightarrow RGB} = (u_{IR \rightarrow RGB}, v_{IR \rightarrow RGB}) = \pi(P_{\Delta s}[u_{IR}, v_{IR}, 1]^T) \quad (7)$$

where $I_{RGB}^s, I_{IR}^{s+\Delta s}$ represent the RGB and IR image at the shooting position s and $s + \Delta s$, respectively; $ROI(\bullet)$ denotes the function to extract the region of interest; $B(\bullet)$ is the binarisation function; $SSIM(\bullet)$ and $PSNR(\bullet)$ represent the SSIM and PSNR functions, respectively; $Q(\Delta s)$ measures the registration quality; $I_{IR \rightarrow RGB}$ is the IR image reprojected into the RGB coordinate, with pixel coordinates $(u_{IR \rightarrow RGB}, v_{IR \rightarrow RGB})$.

After determining the optimal offset Δs , all corresponding IR images are projected into the RGB coordinate frame using Eq. (7). The projected IR images are then fused with their RGB counterparts at a 1:1 ratio to generate fusion images, thereby completing the 2D pixel-level cross-modality registration process. For the 3D-level rendering, the RGB, IR, and fusion layers are rendered by assigning vertex colours sampled from their spatially registered 2D images. Specifically, the vertex colours are designated by sampling pixel values from the corresponding RGB, IR, or fusion images through the established 2D-3D correspondences $C_{2D \leftrightarrow 3D}(p_i^j, X_j)$. During rendering, barycentric interpolation is applied within each triangular face, where fragment colours are weighted according to the barycentric distances of the vertices, thereby reconstructing the MSIF 3D model.

2.2. Defect relocalisation

2.2.1. Experimental setup

In the relocalisation stage (highlighted by the blue box in Fig. 2), GPS data guides engineers to the approximate area of defects. An unregistered image captured via the portable device undergoes feature extraction and matching with images in the pre-inspection database. Based on the stored priors (2D-3D correspondences and SPC), the camera pose of the portable device is estimated by using the Perspective-n-Point (PnP) algorithm (Li et al., 2012; Kongkaew et al., 2022). This enables the projection of multimodal defect layers onto the on-site image through perspective and UV mapping, achieving spatially anchored relocalisation and visualisation of surface defects on the railway track.

In this stage, experiments were conducted by positioning the portable device at varying horizontal distances and vertical heights relative to one edge of the rail to capture on-site images. The reference plane for distance measurement is defined by the rail cross-section and the central point of the device's lower boundary, within which the measurement points on the rail edges and the device are constrained. Based on the ergonomic range for handheld operation, the horizontal distances (h) were set to 20 cm, 40 cm, and 60 cm. The vertical heights (v) were 0.5 m, 1.0 m, and 1.5 m, generating nine shooting positions as illustrated in Fig. 4. An iPhone was utilised as the portable device for image acquisition, mounted on a tripod to ensure accurate positioning during each capture, and producing images at a resolution of 4032×3024 pixels. At each capture position, the iPhone's viewing angle was adjusted to simulate potential variations in inspection viewpoints, while ensuring that the target rail segment remained within the FOV.

2.2.2. 2D-3D correspondences

The on-site camera frame (the unregistered image I_{un} from the same scene) is matched against the pre-inspection image database (with feature point correspondences M_{RGB}) to establish new correspondences ($M'_{RGB}(I_{un})$) concerning the coordinate system defined by the image i , following Eq. (1). Based on new correspondences $M'_{RGB}(I_{un})$, the new 2D-3D correspondences $C_{2D \leftrightarrow 3D}^A(I_{un}, X_j)$ are constructed by querying the previous correspondence set $C_{2D \leftrightarrow 3D}(p_i^j, X_j)$ established during the triangulation process. This process associates each 2D feature point in I_{un} with its corresponding 3D coordinate in the SPC.

2.2.3. Perspective-n-Point (PnP)

The PnP algorithm estimates the intrinsic and extrinsic parameters of the portable camera frame by minimising the reprojection error based on the pinhole camera principle. Given the new 2D-3D correspondences $C_{2D \leftrightarrow 3D}(p_{un}^j, X_j)$, where the j -th 3D points X_j corresponds to the original feature pixel position p_{un}^j , the 3D points are reprojected onto the image plane as p_{un}^j , as expressed

$$p_{un}^j = \pi(K[R_{un}|t_{un}]\tilde{X}_j)K \in \mathbb{R}^{3 \times 3}[[R_{un}|t_{un}]] \in \mathbb{R}^{3 \times 4} \quad (8)$$

$$p_{un}^j = [u_{un}^j, v_{un}^j]^T \tilde{X}_j = [x_j, y_j, z_j, 1]^T$$

where $[R_{un}|t_{un}]$ is the extrinsic matrix of the unregistered image, including rotation (R_{un}) and translate (t_{un}); K is the intrinsic matrix.

The difference between the reprojected feature point p_{un}^j and its original pixel position p_{un}^j is defined as the reprojection error r_{un} , as expressed

$$r_j(K, R_{un}, t_{un}) = p_{un}^j - p_{un}^j \quad (9)$$

To minimise this error, a nonlinear least-squares optimisation is applied to solve the camera parameters (K, R_{un}, t_{un}) of the unregistered images I_{un} . This process, known as bundle adjustment, is also used in the optimisation of a sparse point cloud, and can be written as

$$\underset{K, R_{un}, t_{un}}{\operatorname{argmin}} \sum_j \|r_j(K, R_{un}, t_{un})\|^2 \quad (10)$$

$$s.t. \quad R_{un}^T R_{un} = I_3, \det(R_{un}) = 1$$

$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}, f_x, f_y > 0$$

where f_x, f_y is the focal length of the x and y axes; c_x, c_y is the pixel position of the principal point in the image I_{un} .

According to the projection model (Eq. (8)), the MSIF 3D model is projected onto the on-site camera frame I_{un} based on the estimated pose, which generates the point cloud on the on-site image. Given the estimated pose of the portable camera frame, the MSIF 3D layers are rendered from the same viewpoint within the digital environment, resulting in a synthetic image. This synthetic image is then spatially aligned and overlaid onto the on-site frame, producing the final

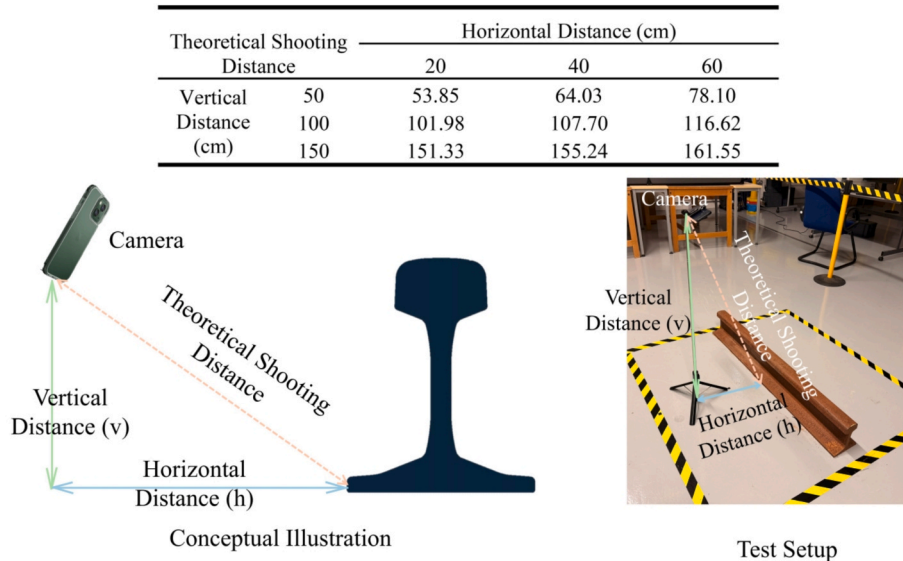


Fig. 4. Illustration of the relocalisation experiment setup. The theoretical shooting distances are the Euclidean distances computed from the vertical and horizontal distances based on the Pythagorean theorem.

augmented view for defect visualisation, as shown in Fig. 2. To handle visibility and simulate occlusion, the rendering process adopts the Painter's Algorithm (Foley, 1996), which draws surfaces from back to front. Algorithm 1 outlines the pseudocode for projecting 3D MSIF defect information onto the on-site image by estimating the portable camera pose through registration with the reconstructed 3D model, thereby generating a spatially registered on-site projected image, as illustrated in Fig. 2.

Algorithm	Relocalisation based on 3D MSIF Model
1	
Input:	Pre-inspection RGB Image Set $I_{RGB} = \{I_{RGB}^1, I_{RGB}^2, \dots, I_{RGB}^n\}$ Pre-inspection IR Image Set $I_{IR} = \{I_{IR}^1, I_{IR}^2, \dots, I_{IR}^n\}$ On-site Image I_{on} Offset $\Delta s = 5$
Output:	Projected On-site Image I_{on}^p
1:	//Step 1: 3D Reconstruction
2:	features, matches = match_feature(extract_feature(I_{RGB})) //Feature extraction & matching
3:	SPC, $C_{2D \leftrightarrow 3D}$ = triangulate(matches) //Get sparse point cloud and 2D-3D correspondences
4:	DPC = CMVS_DMVS(SPC) //Dense Point Cloud generation
5:	mesh = Poisson_reconstruction(DPC) //Mesh generation
6:	//Step 2: Cross-Modality Registration
7:	for $i = 1$ to n - Δs :
8:	T = estimate_transform(match_features($I_{RGB}[i]$, $I_{IR}[i + \Delta s]$))
9:	$I_{IR}[i]$ = apply_transform($I_{IR}[i + \Delta s]$, T)
10:	$I_{fusion}[i]$ = fuse_images($I_{RGB}[i]$, $I_{IR}[i]$)
11:	end for
12:	//Step 3: Relocalisation
13:	F_{on} = extract_feature(I_{on}) // Extract on-site image features
14:	M_{on} = match_feature(F_{on} , features) // Match on-site features with reconstruction features
15:	$C_{2D \leftrightarrow 3D}^{on}$ = build_corr2D3D(F_{on} , $C_{2D \leftrightarrow 3D}$) //Reconstruct 2D–3D correspondences for the on-site image
16:	P_{on} = solvePnP($C_{2D \leftrightarrow 3D}^{on}$) // estimate camera pose via PnP
17:	I_{on}^p = render(mesh, P_{on} , I_{RGB} , I_{IR} , I_{fusion})
18:	return I_{on}^p

2.3. Quantitative evaluation metrics

To quantitatively evaluate the outcomes described in Section 2.2, three-level evaluation is employed progressively to examine the perceptual separability of the projection and the spatial accuracy of defect relocalisation, as shown in Fig. 5. At the first level, the projected perceptual separability of the on-site projected track defect regions is assessed. This is achieved by computing the Edge Strength (Zhang et al.,

2013) of the projected areas using the segmented projection images that retain both projected defect and underlying track surface textures, where the real-world rail textures are treated as background noise. As the second level, the defect-related perceptual separability is evaluated to quantify the visibility of defects within the projection itself by using Contrast-to-Noise Ratio (CNR) (Geissler et al., 2007). During this process, only the projected defects are considered as defective signals, while all other textures in the projection are treated as noise. The final level is to analyse geometric similarity between the projected defects and real-world defects. This involves applying fixed-threshold binarisation to both the projected and real defect regions, followed by calculating their centroid displacement, Intersection-over-Union (IoU), and Structural Similarity Index (SSIM).

3. Results

The relocalisation system renders RGB, IR, and fused projections derived from pre-captured NDT data onto on-site images, enabling spatially aligned relocalisation and visualisation of railway track defects within the bounds of GPS positioning error. At the reconstruction stage, all 233 images were successfully registered to generate the 3D geometric model with a mean reprojection error of 0.66 pixels. On average, 8220 feature points per image were extracted, among which 4712 were successfully associated with the 3D points as valid observations. Compared with a 3D model generated by smartphone LiDAR, the proposed model had an average error of 0.38 mm and a standard deviation of 0.40 mm, as discussed in the previous studies of Wang et al. (2024a,b) and Zhang et al. (2023a,b). At the relocalisation stage, 50 images were captured, of which 42 were successfully registered, corresponding to a registration success rate of 84%. All projections depict both structural and non-structural defects, as well as the adjacent regions. The point-cloud-projected image displays the defective rail segment (as shown in Fig. 6) mentioned in Section 2.1.1 within the portable camera's FOV. The RGB, IR, and Fusion projected images were spatially cropped to retain only the defective regions, as shown in Fig. 7. The RGB projection provides high-fidelity surface textures. The IR projection highlights defect regions with strong thermal contrast, accentuating both their positions and thermal characteristics. The fusion projection combines the thermal saliency of the IR layer with the geometric texture of the RGB layer, mitigating background interference and enhancing defect distinguishability. The results are evaluated qualitatively and quantitatively in the subsequent sections. The qualitative analysis focuses on visual fidelity, clarity of defect representation, and the visibility of occluded regions.

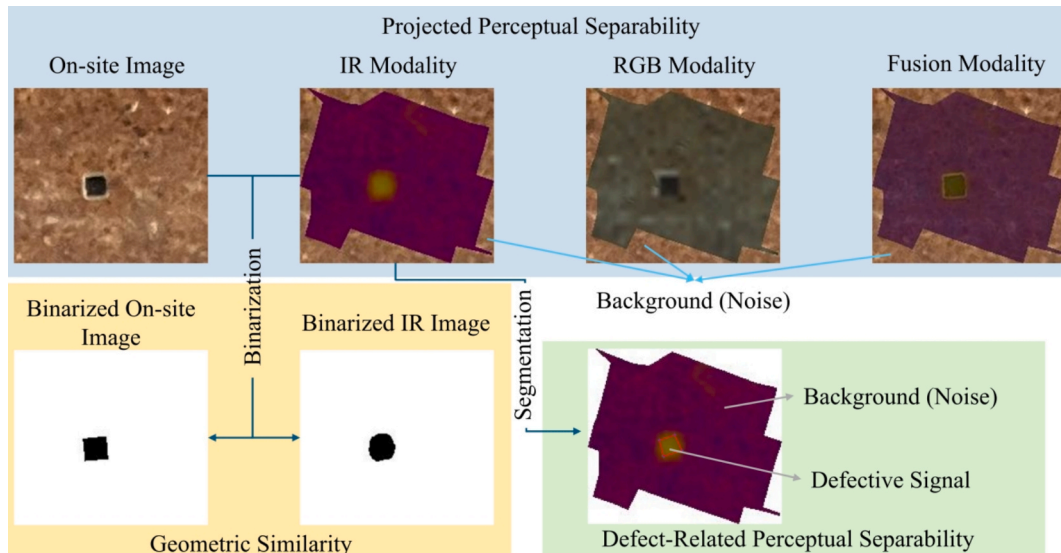


Fig. 5. Illustration of the proposed three-level evaluation.

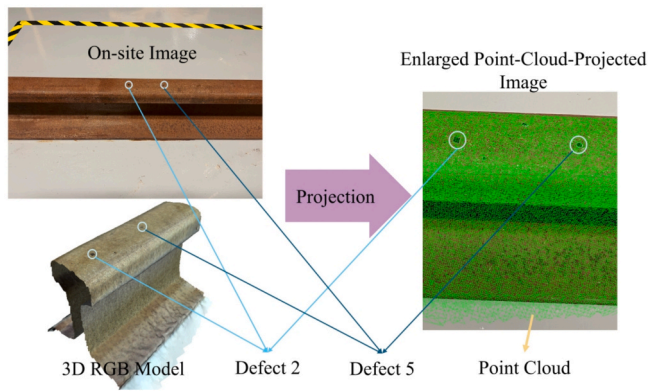


Fig. 6. Illustration of the 3D RGB model and the on-site image with point-cloud projection. The on-site image, shown in top left, was captured at a height of 40 cm and a lateral offset of 50 cm. To avoid occlusion of the rail surface and defects by the projected data, the point-cloud-projected image is enlarged for visualisation. Two identifiable defects (Defects 2 and 5) are highlighted in the figure.

3.1. Qualitative evaluation

From an overall visual perspective, the projected results present spatially aligned and well-rendered overlays across nine shooting positions, with three representative examples shown in Fig. 7(a). To independently investigate the visual characteristics of defects, the images were segmented and enlarged, as illustrated in Fig. 7(b). For each shooting position, non-structural defects (defects 1–5), structural defects (defects 6,7), and the occlusion defect (defect 6) are all correctly visualised. It is observed that background surface details become increasingly blurred with distance; the projected regions, on the other hand, maintain consistent sharpness and contrast, exhibiting negligible perceptual variation to the human eye. Based on the segmented defect regions in Fig. 7(b), the edges of the projected areas become less distinct with increasing distance. However, this degradation remains minimal within the projected areas themselves. The sustained clarity is attributed to the use of vector-based rendering, which preserves spatial fidelity and high-contrast contours. Additionally, in partially occluded areas, such as defect 6 (head-web junction), the projected defects may not be evident due to geometric occlusion. In contrast, their spatial alignment within the projected scene remains unaffected, as shown in Fig. 7(b). The painter-based surface reconstruction algorithm handles these effects. It does not affect the spatial localisation of defects, since occluded regions

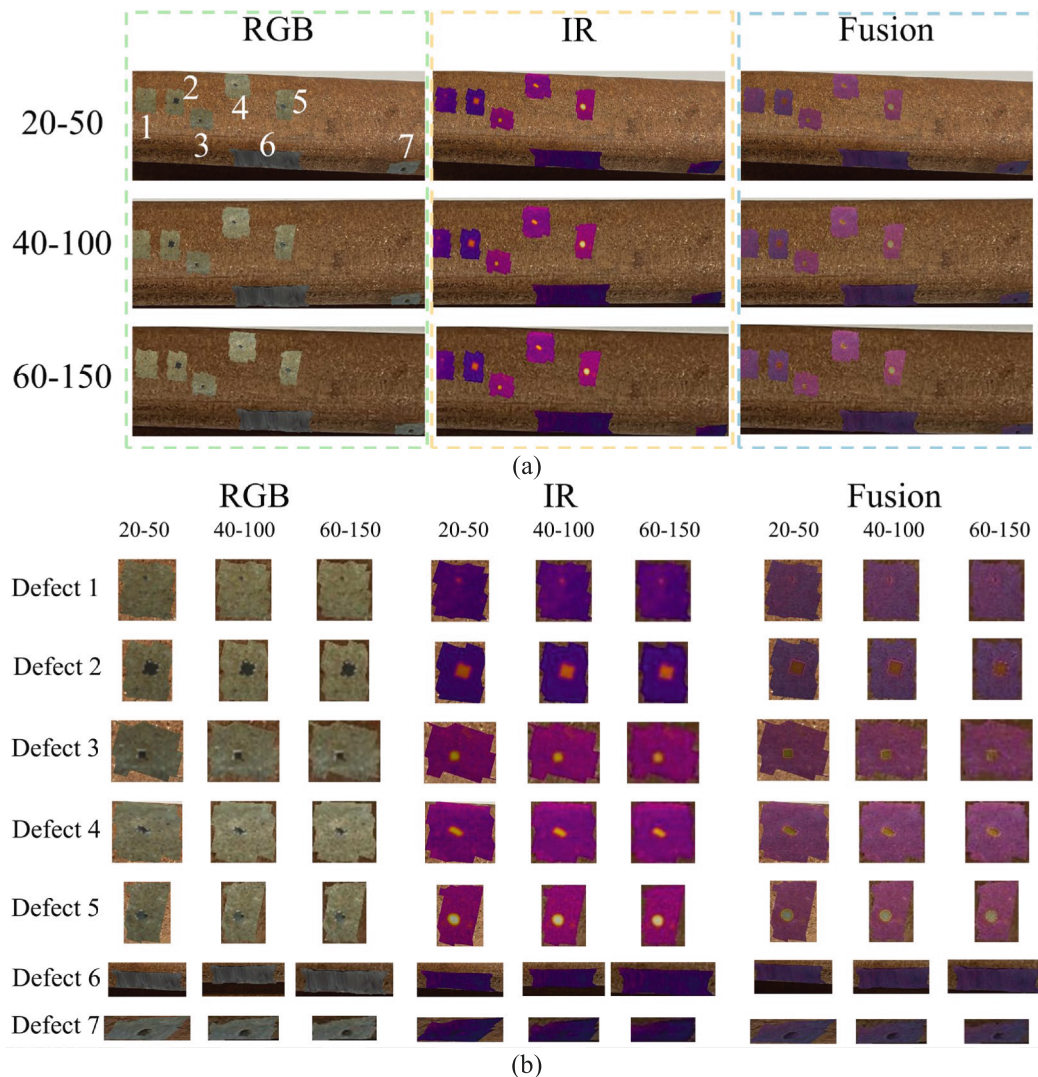


Fig. 7. Qualitative relocalisation results. (a) Projections of RGB, IR, and fused modalities on on-site images captured at three viewing distances; (b) Enlarged defect regions with background context. Despite differing colour sources, all modalities show consistent projection positions and sizes.

can become visible when the on-site view is taken from alternative viewpoints. This particular defect is considered only for the qualitative evaluation of occlusion and is excluded from the subsequent quantitative analysis.

Due to the presence of an external light source, the projected regions in the RGB modality exhibit a noticeable colour deviation from their original appearance. The geometric edges of the defects are sharply outlined in the projected RGB images, facilitating sharp visual separation. However, minor defects, such as defect 1, display weak contrast due to their limited size. In contrast, the IR projected images display a red-tinted projection that markedly deviates from the original track surface colour. For the non-structural defects (defects 1–5), the defective areas are highlighted in yellow, exhibiting a high contrast relative to the surrounding non-defective regions. However, the structural defects (defects 6 and 7) appear with blurred boundaries in the IR modality compared to the RGB modality. This is because thermography captures surface temperature distributions, whereas RGB imaging is sensitive to variations in visible light. Structural defects typically alter surface reflectivity or emissivity slightly, resulting in minimal thermal variation. In contrast, non-structural defects alter the surface emissivity, resulting in discernible thermal gradients under external thermal excitation. Combining RGB and IR modalities, the fusion projected images preserve the geometric textures from the RGB projected images and the contrast between defective and non-defective regions from the IR projected images. The resulting visualisations effectively represent both structural and non-structural defects, as well as small-scale defects. However, due to the 1:1 fusion ratio, both texture fidelity and contrast of defects are decreased, resulting in a visually weaker appearance compared with the individual RGB and IR modalities.

3.2. Projected perceptual separability between projection and track background

The perceptual separability between the projected area and the original-colour track background was evaluated using Edge Strength (ES), as shown in Fig. 8. For each shooting distance, the evaluation was focused on the segmented defect regions shown in Fig. 7(b). Across all shooting distances, the positive ES values indicate that the projected areas consistently exhibit higher edge strength than the background, enabling a clear visual distinction from the track surface.

As shown in Table 2, the IR projected regions exhibit the highest ES, suggesting the strongest visual separability from the background, followed by the RGB and fusion projections. This observation is consistent

Table 2

Average edge strength (ES) and Contrast-to-noise ratio (CNR) across different modalities.

Indicators	RGB	IR	Fusion
Average ES	0.0156	0.0171	0.0123
Average CNR	1.253	1.144	0.299

with the qualitative evaluation. The superior performance of the IR modality arises because thermal information is independent of surface colour, allowing the projections to remain unaffected by background texture. In comparison, the RGB projection benefits from external illumination, which improves the brightness of the projected region. However, its colour remains partially influenced by the colour and texture of the track surface, reducing separability. As a derived modality, the fusion projection integrates texture (as noise for separability) and the reduced thermal contrast by averaging the IR pixel values, resulting in the lowest ES among the three modalities.

For the structural defects, the projected ES values are higher than those for structural defects, as shown in Fig. 9. However, this does not imply that RGB projections are more distinguishable from the projection background than IR projections. This is because the ES metric is computed based on the global edge gradient. The edges of the structural defects appear sharper in the RGB-projected images than in the IR-projected ones. Therefore, this observation does not contradict the earlier conclusion that IR-projected regions remain the most visually distinguishable from the projection background.

3.3. Projected perceptual separability of defects within the projected regions

This analysis examines the influence of modalities and shooting distances on the perceptual separability of defects within the projected regions, excluding background. As illustrated in Fig. 10(b) (except defect 6), the defect areas within the projections are defined as target regions, while all remaining areas are treated as noise. For each modality and shooting distance, the perceptual separability of defects is quantitatively evaluated using the contrast-to-noise ratio (CNR), as shown in Fig. 10(a).

As expected, the CNR values of both the IR and RGB modalities are higher than those of the fusion modality. With increasing shooting distance, CNR decreases across all modalities, indicating that the defects become progressively less distinguishable. This observation is consistent with the earlier qualitative analysis.

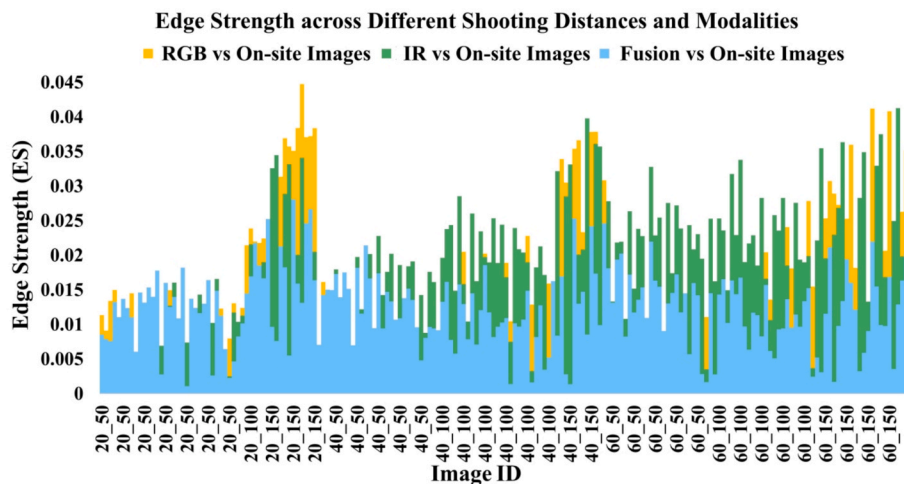


Fig. 8. Edge Strength bar chart for all defects concerning nine shooting positions and three modalities (including RGB, IR and Fusion modalities). The Image ID is denoted as “h_v”, where h and v represent the horizontal and vertical distances, respectively. The x-axis distribution is non-uniform, as several defects fall outside the field of view.

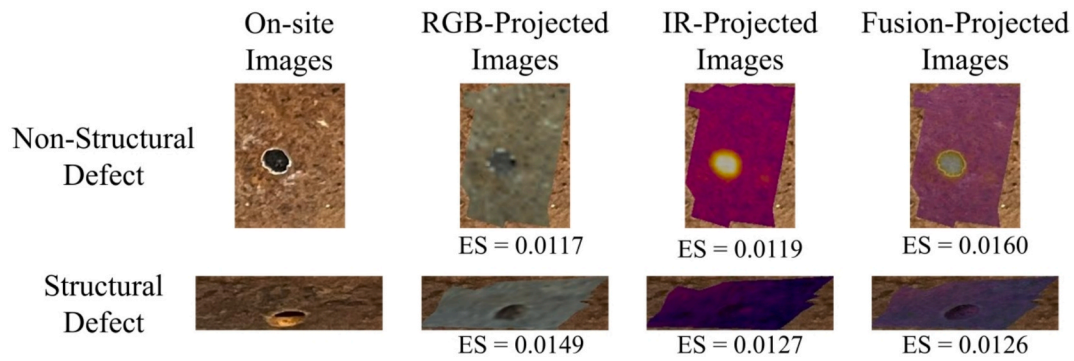


Fig. 9. ES values and images of structural and non-structural defects across original AR, RGB-projected, IR-projected, and fusion-projected images.

Unlike the ES results, the CNR values of the RGB modality are slightly higher than those of the IR modality (shown in Fig. 10(b) and Table 2) and exhibit a less pronounced decline with increasing shooting distance. This difference is attributed to the spatial alignment between the projected defects and the on-site image defects. Although the defect contours in the RGB projections become somewhat blurred at greater distances, they generally remain confined within the labelled defect regions. In contrast, the defect areas observed in the IR modality tend to exceed the actual defect boundaries due to thermal diffusion effects. This phenomenon is also clearly reflected in the fusion projections. As a result, the overflowed thermal regions in the IR projections are considered as noise, thereby reducing the overall CNR.

Additionally, CNR is sensitive to variations in pixel intensity within the target region. Structural defects in the RGB modality typically exhibit sharper boundaries and higher local contrast between defective and non-defective areas, which elevates CNR values in the RGB projections (Fig. 11). Moreover, as shooting distance increases, the number of pixels involved in the calculation decreases, which in turn amplifies the influence of individual pixels on the CNR value. The inclusion of thermally diffused pixels outside the annotated defect area in the IR modality further lowers the CNR at greater distances. This explains why the CNR of the RGB modality declines more slowly than that of the IR modality.

3.4. Spatial and structural fidelity evaluation

Given the high perceptual separability of the projected areas and corresponding defects in the defect-projected images, this section focuses on evaluating spatial and structural fidelity. The binarised defect regions from Fig. 7(b) are used to assess the spatial consistency and visual fidelity of the projected defects across different shooting distances. Since all modality projections are generated using the same transformation matrix, the projected defect locations are considered spatially consistent across modalities. As a result, the specific influence of modality or defect type is not the focus of this section and will not be further discussed.

3.4.1. Spatial consistency – centroid error

Fig. 12(a) quantifies the Euclidean offset between projected and real-world defect positions. The centroid error increases with both vertical (v) and horizontal (h) camera distances, rising from 0.34 mm at $h = 0.2$ m and $v = 0.5$ m to 3.25 mm at $h = 0.6$ m and $v = 1.5$ m. Notably, although the error increases with distance, the marginal rise diminishes. Especially, from $(h = 0.4$ m, $v = 1.5$ m) to $(h = 0.6$ m, $v = 1.5$ m), the error even decreases slightly. However, the centroid error at $(h = 0.2$ m, $v = 0.5$ m) is the minimum due to its high clarity and high-contrast texture.

3.4.2. Visual fidelity – IoU

The IoU between projected and actual defect regions quantifies their

effective overlap, as shown in Fig. 12(b). As the camera working distance increases, the IoU decreases from 0.993 at $(h = 0.2$ m, $v = 0.5$ m) to 0.961 at $(h = 0.6$ m, $v = 1.5$ m), reflecting reduced matching accuracy at larger shooting distances. There is also a minor marginal degradation and a slight rebound from 0.959 at $(h = 0.4$ m, $v = 1.5$ m) to 0.961 at $(h = 0.6$ m, $v = 1.5$ m).

3.4.3. Visual fidelity – SSIM

SSIM evaluates the perceptual similarity between projected and actual defect regions by comparing contrast (local standard deviation) and structural information (covariance) within binarised image blocks. As shown in Fig. 12(c), SSIM decreases with increasing camera distance, dropping from 0.968 at $(h = 0.2$ m, $v = 0.5$ m) to 0.845 at $(h = 0.6$ m, $v = 1.5$ m). Although SSIM reaches its minimum of 0.845 at the most significant distance, the marginal decrement is decreased.

3.5. Working distances

Based on the above quantitative and qualitative analyses, the performance of the projected images deteriorates as the shooting distance increases. This deterioration can be attributed to two factors: an increasingly orthographic viewing geometry, and the smaller pixel footprint of defect regions. Under such conditions, scenes approach an orthographic projection, leading to reduced parallax and consequently weaker geometric constraints for pose estimation. As the shooting distance increases, the effective spatial resolution decreases, resulting in fewer detectable features and less reliable feature correspondence, thereby increasing errors in pose estimation.

However, the rate of decline in IoU becomes less pronounced at larger shooting distances. This effect arises because defect regions span fewer image pixels at greater distances, rendering pose estimation less sensitive to pixel-level noise and mismatches filtered by RANSAC.

4. Discussions

It should be noted that this research focuses on a proof-of-concept for relocalisation at the pixel and feature levels, rather than at the decision level. Specifically, it addresses the challenges of surface defect relocalisation in the existing workflows, where spatial anchors between NDT data and GPS-derived uncertainty zones are typically absent. As such, real-time performance and parameter sensitivity analyses are beyond the scope of this work. The proposed system should be regarded as an early-stage exploration of AR-assisted relocalisation for railway inspection. Compared with RTK-assisted (Mikhaylov et al., 2023) and UWB-assisted systems (R. Wang et al., 2024), the proposed method requires further experimental validation in terms of maturity and automation. While it currently lacks the capability for continuous operation, the system possesses strong theoretical potential as an infrastructure-free solution that does not rely on external positioning hardware. In terms of quantitative performance, the mean relocalisation error

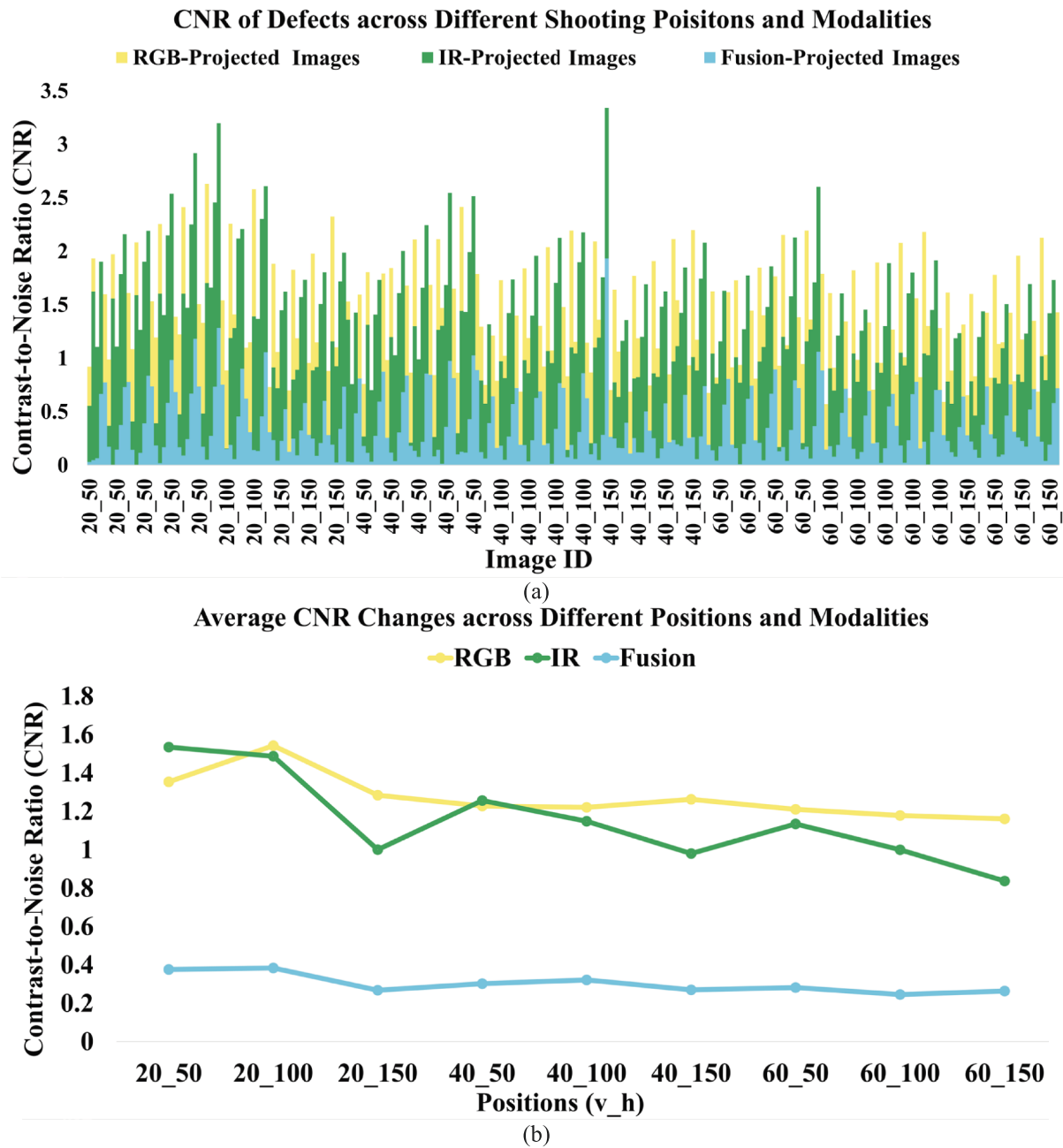


Fig. 10. CNR bar chart (a) and average CNR bar chart (b) for all defects concerning nine shooting positions and three modalities. The Image ID is denoted as “h_v”, where h and v represent the horizontal and vertical distances, respectively. For the CNR bar chart in (a), data from incompletely projected defects have been excluded to ensure measurement validity, causing the x-axis non-uniform.

observed in this study ranges from 0.34 to 3.62 mm, which is comparable to existing AR-based systems (Al-Sabbag et al., 2022; Chi et al., 2022), particularly considering the challenging texture of railway tracks. For example, although Al-Sabbag et al. (2022) reported a slightly lower average error of 0.41 mm in AR-guided wall inspection, their target surfaces contain abundant and distinct features. In contrast, railway track surfaces in this study exhibit sparse discriminative features and highly repetitive patterns, substantially increasing relocalisation difficulty. Similarly, a Vuforia-based structural inspection system for reinforced concrete achieved an average positioning error of 1.83 mm (Chi et al., 2022). However, that approach relies on additional sensing equipment, such as laser scanners, to obtain object position.

4.1. 3D MSIF reconstruction

As demonstrated in Wang et al. (2025), the 3D MSIF reconstruction workflow employed in this paper achieves millimeter-level geometric accuracy while simultaneously enabling 3D multimodal fusion. Given the focus on maintenance workflow, this study did not optimise the active thermography process, including the selection of the excitation source, heating duration, or IR camera settings. Moreover, the artificial defects used in this study were designed to support the evaluation of relocalisation accuracy, rather than to replicate the full range of real-world track degradation, damage, or defect types. As reported in Chen et al. (2021), under sufficient thermal excitation, structural defects, such as defects 6 and 7 in this study, can exhibit enhanced thermal contrast and reveal subsurface defects in the IR imagery.

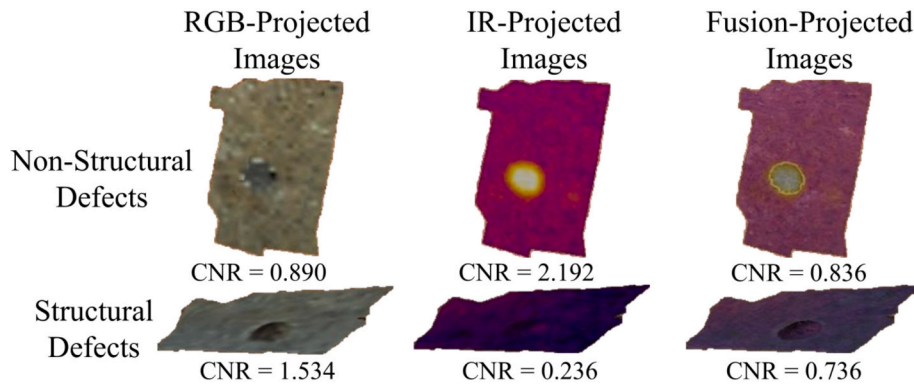


Fig. 11. CNR and images of structural and non-structural defects across original AR, RGB-projected, IR-projected, and fusion-projected images.

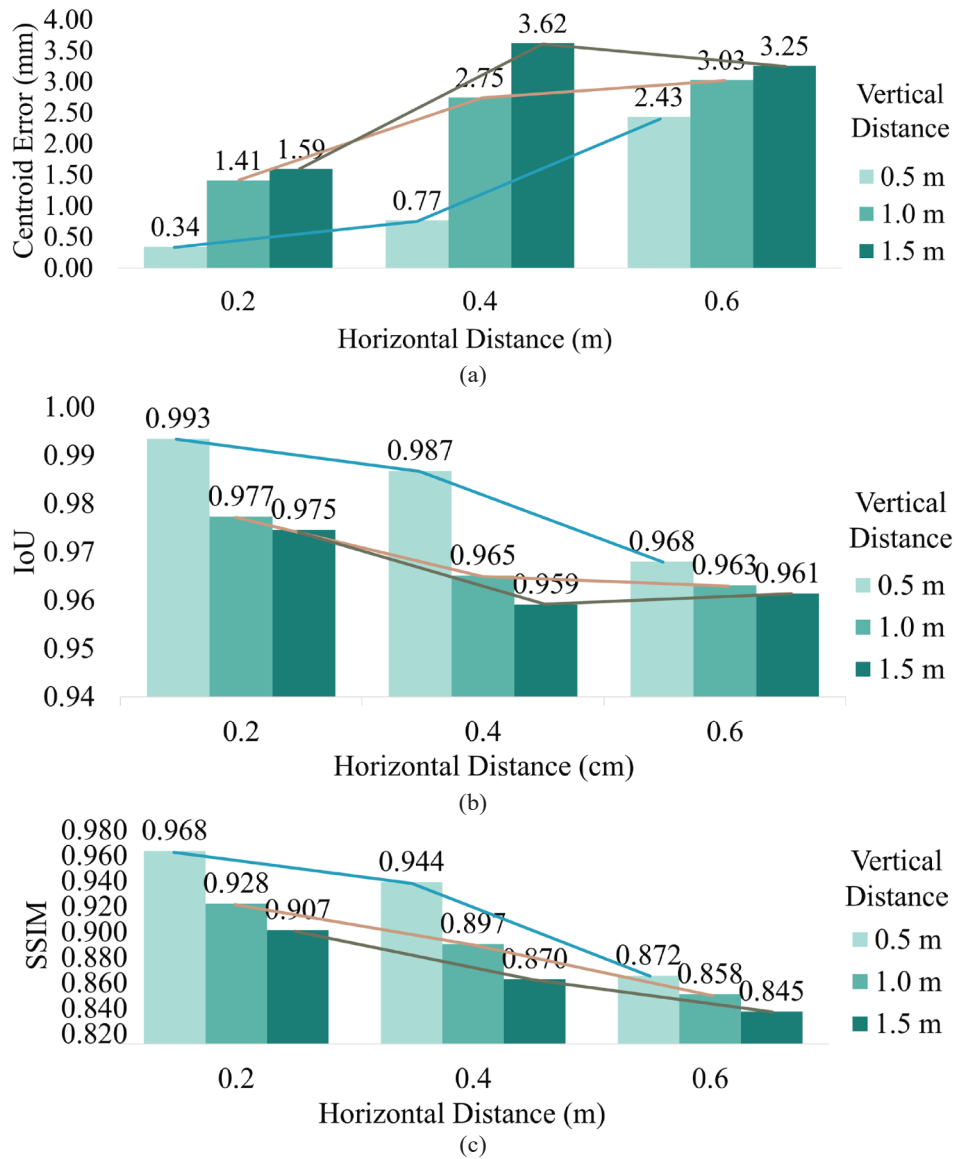


Fig. 12. Combined plot of (a) centroid error, (b) IoU, and (c) SSIM between projected and actual defect regions with respect to horizontal and vertical distances.

Unlike the simultaneous seven-camera acquisition setup proposed in Wang et al. (2025), this study employed a sequential acquisition strategy due to hardware constraints. Under stable indoor lighting conditions, the sequential acquisition maintained the same viewing geometry as the

seven-camera setup, ensuring consistent track positioning within the scene. Nevertheless, both numerical and spatial temperature drifts were observed. The numerical drift arises from repeated activation of the thermal excitation source, which affects IR frames acquired at later

stages. Because of the distinct heat absorptivity between defective and non-defective areas, the temperature drift (around 0.2 °C) of non-defective areas is significantly smaller than the temperature difference (around 2 °C) between defective and non-defective areas. Although this drift does not affect the final inspection conclusions, the resulting temperature distribution of all railway tracks is slightly different from the idle condition, representing a systematic error introduced by the data acquisition process. Regarding spatial drift, the use of a fixed global transformation matrix leads to pixel-level misalignment between RGB and IR, particularly towards the image periphery, while the inspection outcomes are also not affected. To address the above two shifts, future work will focus on thermal compensation methods (Restrepo-Girón and Loaiza-Correa, 2016; Li et al., 2023) and machine learning-based approaches (Lielāmuris et al., 2023; Moscovini et al., 2024) to correct temperature bias and reduce local spatial misalignment. Additionally, the degree of automation is limited by manual operations, such as denoising of the point clouds and mesh, which may be solved by introducing deep learning methods (Shen et al., 2022; Wang et al., 2023).

To enable the practical deployment, the data acquisition system should be upgraded to support synchronous multi-angle capture along a common horizontal plane, ensuring both data completeness and acquisition efficiency. In particular, the active thermography module should adopt thermal excitation sources with high power density and rapid response to enhance the contrast of surface and subsurface defects in IR imagery under varying environmental and dynamic conditions.

4.2. Relocalisation stage

Although our workflow eliminates the need for on-site reinspection and provides precise, multidimensional visualisations of defects, the 3D MSIF reconstruction depends on manual image registration and operator-guided mesh trimming. This offline preparation is generally acceptable in practice, as rail maintenance typically follows defect detection after a delay. The severity of the damage determines the timing of intervention. According to the Federal Railroad Administration (Office of Railroad Safety, 2015), A, A2, and B-level defects (compound fissures, transverse fissures, detail fractures, and engine burn fractures (Class A); bolt-hole cracks, horizontal split heads, and vertical split heads (Class A2); and broken bases, head-and-web separations, split webs, and piped rails (Class B) requires immediate verification. All other defect indications must be verified within at least 4 h. This regulatory framework implies that the workflow is suitable for the vast majority of surface defects and can support both initial verification and subsequent maintenance activities, such as rail replacement and rail grinding.

Although this study demonstrates the feasibility of integrating multimodal data across different stages of the railway track maintenance workflow, practical industrial applications demand real-time performance. Future work will therefore focus on model parameter optimisation, the adoption of lightweight algorithms, and the selection of appropriate feature point extraction and matching strategies. These efforts may benefit from existing AR-related algorithms and applications, such as markerless real-time algorithms (Zhou et al., 2023), transformer-based convolutional neural networks (Chindiyababy et al., 2025), and Auto Anchoring Systems (Mohammadkhorasani et al., 2023). In addition, feature extraction and matching algorithms should be systematically evaluated and optimised for railway inspection tasks to identify parameter configurations that achieve an appropriate trade-off between acquisition speed and relocalisation accuracy.

User interaction and ergonomics were not directly explored in this study, but they represent critical factors for field deployment. Poor interface layout or complex control schemes can increase operator learning time and even induce simulator sickness. In future work, the user interface should undergo human factors evaluations, focusing on ergonomic design and workload reduction, to safeguard operator health

and improve usability during prolonged field operations. Furthermore, embedding preloaded spatial maps with defined operational boundaries into portable devices may help constrain video data acquisition to authorised areas, thereby supporting the preservation of privacy and compliance with data governance policies.

4.3. Outdoor dynamic deployment

In contrast to the controlled indoor environment used in this study, outdoor deployment introduces various uncertainties, including variable lighting conditions and occlusions during dynamic inspections. Intense lighting can cause specular reflections and overexposed regions, while low-light conditions reduce image contrast and hinder the reliable feature extraction from RGB images. Occlusions further reduce the number of valid features and introduce outlier correspondences. These challenges could be mitigated by applying deep learning-based feature extraction and matching methods, such as SuperPoint (Fan et al., 2025) and LoFTR (Rothenberger et al., 2025). Motion blur, induced by hand-held operation or rapid relative movement between the acquisition system and the track during NDT data capture, serves as another critical source of image degradation. As both the reconstruction and relocalisation modules rely heavily on accurate visual feature matching, motion blur can lead to unstable correspondences, increased registration errors, and even complete relocalisation failure. From an algorithm perspective, deep learning-based image enhancement methods (Chen and Yu, 2024) and cross-modal attention mechanisms (Sun et al., 2022) have great potential to enhance image quality of railway tracks. Furthermore, the proposed acquisition system can be combined with integrated global shutter cameras and IMU-assisted deblurring techniques (Arslan et al., 2024) to further mitigate motion blurring.

To improve overall robustness and ensure the pipeline's generalisability across diverse environmental conditions and real-world defect types, future work should focus on conducting extensive outdoor experiments and establishing representative datasets. Additionally, optimising reconstruction algorithms and upgrading processing platforms to minimise latency and mitigate motion blur will be crucial for developing a real-time, automated, and multimodal 3D inspection system that meets the rigorous demands of practical railway maintenance.

5. Conclusion

This research proposes a defect maintenance workflow for mapping surface defects on railway tracks by overlaying 3D multimodality sensor information fusion (MSIF) non-destructive testing (NDT) data onto on-site scenes. This workflow establishes the relative projection between the on-site image and 3D multi-sensor model via the perspective-n-point (PnP) algorithm. As a result, in the GPS-derived uncertainty area, the track defects can be relocated with spatially aligned visualisation of pre-inspection signals. The projected meshes in the RGB and IR modalities are visually distinguishable from the on-site background, achieving an average edge strength (ES) of 0.0156 and 0.0171 correspondingly. In addition, the defects within the projected region are highlighted with an average contrast-to-noise ratio (CNR) of 1.253 (RGB) and 1.144 (IR), indicating strong perceptual separability between defective and non-defective areas in both modalities. The relocalisation stage achieves defect centroid offsets ranging from 0.34 mm to 3.62 mm under hand-held operation.

For visualisation, RGB projection delivers high-fidelity surface textures and uncovers defects at the head-web junction that are often hidden under oblique viewing angles. Infrared projection emphasises thermally salient regions, enhancing both defect localisation and thermal signature. Fused projection combines geometric detail with thermal contrast to suppress background noise and to improve defect discernibility. Quantitative validation demonstrates Intersection-over-Union (IoU) values ranging from 0.956 to 0.993 and Structural Similarity Index Measure (SSIM) values ranging from 0.845 to 0.968,

confirming that the proposed workflow can maintain high spatial precision and visual fidelity under controlled conditions.

This research demonstrates the potential of digital technologies as a cost-effective and efficient solution for defect localisation, contributing to improving safety and operational efficiency in railway infrastructure. Future work will focus on intuitive interface design, low-latency rendering pipelines optimised for wearable hardware, secure data handling protocols, and end-to-end automation to support high-throughput, real-time railway track inspection.

CRedit authorship contribution statement

Yizhong Wang: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lichao Yang:** Writing – review & editing. **Isidro D. Cardenas:** . **Yifan Zhao:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Ahmad, W., Mutz, M., Werth, D., 2024. Digital twin of rail for defect analysis. In: ACM International Conference Proceeding Series, pp. 53–60. <https://doi.org/10.1145/3657547.3657549>.
- Al-Sabbag, Z.A., Yeum, C.M., Narasimhan, S., 2021. Interactive defect quantification through extended reality. *Adv. Eng. Inform.* 51. <https://doi.org/10.1016/j.aei.2021.101473>.
- Andrianova, E.G., Demidova, L.A., 2021. An Approach to Image Matching Based on SIFT and ORB Algorithms. In: 2021 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). IEEE, pp. 534–539. <https://doi.org/10.1109/SUMMA53307.2021.9632214>.
- Arslan, A., Gultekin, G.K., Saranlı, A., 2024. IMU-aided adaptive mesh-grid based video motion deblurring. *PeerJ Computer Science* 10. <https://doi.org/10.7717/peerj-cs.2540>.
- Bakır, Ç.N., et al., 2023. Use of augmented reality in mental health-related conditions: a systematic review. *Digital Health* 9. <https://doi.org/10.1177/20552076231203649>.
- Bianco, G., et al., 2013. A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. *Sensors* 13 (8), 11007–11031. <https://doi.org/10.3390/s130811007>.
- Bona, J., et al., 2022. Semantic integration of multi-modal data and derived neuroimaging results using the platform for imaging in precision medicine (PRISM) in the arkansas imaging enterprise system (ARIES). *Front. Artif. Intell.* 4, 1–10. <https://doi.org/10.3389/frai.2021.649970>.
- Bruno, R.R., et al., 2022. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. *Critical Care* 26 (1), 1–13. <https://doi.org/10.1186/s13054-022-04202-x>.
- Chen, C.-I. et al. (2009) "Uniscale multi-view registration using double dog-leg method," in M.I. Miga and K.H. Wong (eds.), p. 72611F. Doi: 10.1117/12.810966.
- Chen, K., Yu, L., 2024. Motion deblur by learning residual from events. *IEEE Trans. Multimedia* 26, 6632–6647. <https://doi.org/10.1109/TMM.2024.3355630>.
- Chen, X., et al., 2021. Tomographic reconstruction of rolling contact fatigues in rails using 3D eddy current pulsed thermography. *IEEE Sensors Journal* 21 (17), 18488–18496. <https://doi.org/10.1109/JSEN.2021.3086307>.
- Chen, X., et al., 2022. Automatic detection system with 3D scanning and robot technology for detecting surface dimension of the track slabs. *Automation in Construction* 142. <https://doi.org/10.1016/j.autcon.2022.104525>.
- Chi, H.L., et al., 2022. Rebar inspection integrating augmented reality and laser scanning. *Automation in Construction* 136. <https://doi.org/10.1016/j.autcon.2022.104183>.
- Chindiyababy, U., et al., 2025. Real-Time Interaction in AR/VR Environments: A Deep Learning Approach to Human-Computer Interaction. In: 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT). IEEE, pp. 1253–1258. <https://doi.org/10.1109/CE2CT64011.2025.10939687>.
- Cruz-Mota, J., et al., 2012. Scale invariant feature transform on the sphere: theory and applications. *International Journal of Computer Vision* 98 (2), 217–241. <https://doi.org/10.1007/s11263-011-0505-4>.
- Cucci, D.A., Rehak, M., Skalous, J., 2017. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 1–12. <https://doi.org/10.1016/j.isprsjprs.2017.05.008>.
- Cui, A., Liang, M., 2023. Applying Civil Information Modeling And Augmented Reality To The Construction Of Underground Pipelines. In: in 2023 Winter Simulation Conference (WSC). IEEE, pp. 2734–2743. <https://doi.org/10.1109/WSC60868.2023.10408142>.
- Dargan, S., et al., 2023. Augmented reality : a comprehensive review. *Arch. Comput. Methods Eng.* 30 (2), 1057–1080. <https://doi.org/10.1007/s11831-022-09831-7>.
- Derpanis, K.G., 2010. Overview of the RANSAC Algorithm. *Image Rochester NY* 4 (1), 2–3.
- Engels, C., Stewénius, H., Nistér, D., 2006. Bundle adjustment rules. *Photogrammetric Computer Vision* 2 (32).
- European, G.S.A., 2019. In: Report on Aviation User Needs and Requirements, p. 66.
- Fan, K., Cui, H., Liu, M., 2025. Research on low-illumination image feature extraction algorithm based on improved SuperPoint convolutional neural network. In: Liu, Y. (Ed.), *International Conference on Physics, Photonics, and Optical Engineering (ICPPOE 2024)*. SPIE, p. 1. <https://doi.org/10.1117/12.3056492>.
- Foley, J.D., 1996. *Computer Graphics: Principles and Practice*. Edited by Addison-Wesley Professional.
- Furukawa, Y., et al., 2010. Towards internet-scale multi-view stereo. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1434–1441. <https://doi.org/10.1109/CVPR.2010.5539802>.
- Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (8), 1362–1376. <https://doi.org/10.1109/TPAMI.2009.161>.
- Gabara, G., Sawicki, P., 2018. A new approach for inspection of selected geometric parameters of a railway track using image-based point clouds. *Sensors* 18 (3), 791. <https://doi.org/10.3390/s18030791>.
- Geissler, A., et al., 2007. Contrast-to-noise ratio (CNR) as a quality parameter in fMRI. *J. Magnet. Resonance Imag.* 25 (6), 1263–1270. <https://doi.org/10.1002/jmri.20935>.
- Giulietti, N., et al., 2025. Combined use of infrared imaging and deep-learning techniques for real-time temperature measurement of train braking components. *IEEE Trans. Instrum. Measure.* 74, 1–8. <https://doi.org/10.1109/TIM.2025.3542884>.
- Hartley, R.L., 1997. In: *Defense of the Eight-Point Algorithm*, pp. 580–593.
- Hesch, J.A., Roumeliotis, S.I., 2011. A Direct Least-Squares (DLS) method for PnP. In: 2011 International Conference on Computer Vision. IEEE, pp. 383–390. <https://doi.org/10.1109/ICCV.2011.6126266>.
- Hou, Y., et al., 2024. Global streetcapes — a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS J. Photogram. Remote Sens.* 215, 216–238. <https://doi.org/10.1016/j.isprsjprs.2024.06.023>.
- Imre, E., Hilton, A., 2015. Order statistics of RANSAC and their practical application. *Int. J. Comp. Vision* 111 (3), 276–297. <https://doi.org/10.1007/s11263-014-0745-1>.
- Michael, K., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. *Eurographics Symposium on Geometry Processing*.
- Keylin, A. (2019) "Measurement and Characterization of Track Geometry Data: Literature Review and Recommendations for Processing FRA ATIP Program Data," (August). Doi: <https://rosap.nrl.bts.gov/view/dot/68613>.
- Kongkaew, S., Ruchanurucks, M., Takamatsu, J., 2022. "Automatically landing an unmanned aerial vehicle using perspective- n -point algorithm based on known runway image: area localization and feature enhancement with time consumption reduction. *ASME Open J. Eng.* 1. <https://doi.org/10.1115/1.4055081>.
- Kooban, F. et al. (2023) "Advanced technology in railway track monitoring using the GPR Technique: A Review," (2013).
- Koohmishi, M., Connolly, D.P., 2025. Railway ballast fouling detection using thermal imaging: integration of LSTM and XGBoost. *Trans. Geotech.*, 101889 <https://doi.org/10.1016/j.trgeo.2025.101889>.
- Li, Q. et al. (2023) "Structure-guided camera localization for indoor environments," 202 (June), pp. 219–229.
- Li, S., Xu, C., Xie, M., 2012. A robust o(n) solution to the perspective-n-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (7), 1444–1450. <https://doi.org/10.1109/TPAMI.2012.41>.
- Li, Y., et al., 2023. Adaptive compensation method for the infrared temperature measurement error based on 3-D thermal imaging. *IEEE Sens. J.* 23 (10), 10525–10537. <https://doi.org/10.1109/JSEN.2023.3263902>.
- Lielamurs, E., et al., 2023. Infrared image pre-processing and IR/RGB registration with FPGA implementation. *Electronics* 12 (4), 882. <https://doi.org/10.3390/electronics12040882>.
- Lindeberg, T. (2012) "Scale Invariant Feature Transform," *Scholarpedia*, 7(5), p. 10491. Doi: 10.4249/scholarpedia.10491.
- Liu, K., et al., 2024a. Visual localization and quantitative detection method for thermal defects in cable terminals of high-speed trains based on a temperature derivative. *Measur. Sci. Technol.* 35 (11). <https://doi.org/10.1088/1361-6501/ad6701>.
- Liu, L., et al., 2024b. Fast Rail Surface Inspection Method and System Based on 3D Line Laser Imaging. In: 2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). IEEE, pp. 1–6. <https://doi.org/10.1109/M2VIP62491.2024.10746201>.
- Liu, X., et al., 2025. Detection of rail defects using nondestructive testing methods. In: *Resilient, Sustainable and Smart Ballasted Railway Track*. Elsevier, pp. 249–279. <https://doi.org/10.1016/B978-0-443-33368-2.00021-1>.
- Liu, X., Barkan, C.P.L., Saat, M.R., 2011. Analysis of derailments by accident cause: Evaluating railroad track upgrades to reduce transportation risk. *Trans. Res. Record* 2261, 178–185. <https://doi.org/10.3141/2261-21>.
- Liu, Y.Z., et al., 2022. A novel abnormal detection method for bearing temperature based on spatiotemporal fusion. *Proc. Inst. Mech. Eng. Part F: J. Rail Rapid Transit* 236 (3), 317–333. <https://doi.org/10.1177/09544097211022105>.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* 60 (2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.

- Mahmoud, A., Mohamed, M.G., El Shazly, A., 2022. Low-cost framework for 3D reconstruction and track detection of the railway network using video data. *Egy. J. Remote Sens. Space Sci.* 25 (4), 1001–1012. <https://doi.org/10.1016/j.ejrs.2022.11.001>.
- Mikhaylov, D., et al., 2023. Toward the future generation of railway localization exploiting RTK and GNSS. *IEEE Trans. Instrum. Measur.* 72, 1–10. <https://doi.org/10.1109/TIM.2023.3272048>.
- Mikusova, N., et al., 2024a. Use of augmented reality in railway transport. *Trans. Res. Procedia* 77 (2023), 253–259. <https://doi.org/10.1016/j.trpro.2024.01.034>.
- Mikusova, N., et al., 2024b. Use of augmented reality in railway transport. *Trans. Res. Procedia* 77 (2023), 253–259. <https://doi.org/10.1016/j.trpro.2024.01.034>.
- Miles, Z., et al., 2024. Hybrid multi-modal NDE sensing system for in-motion detection and localization of rolling contact fatigue damage in rails. *NDT & E International* 147. <https://doi.org/10.1016/j.ndteint.2024.103209>.
- Mohammadkhorasani, A., et al., 2023. Augmented reality-computer vision combination for automatic fatigue crack detection and localization. *Comp. Indust.* 149. <https://doi.org/10.1016/j.compind.2023.103936>.
- Mordia, R., Verma, A.K., 2025. Nondestructive testing methods for rail defects detection. *High-speed Railway* 3 (2), 163–173. <https://doi.org/10.1016/j.hspr.2025.03.001>.
- Moscovini, L., et al., 2024. An open-source machine-learning application for predicting pixel-to-pixel NDVI regression from RGB calibrated images. *Comp. Electr. Agri.* 216. <https://doi.org/10.1016/j.compag.2023.108536>.
- Mussatayev, M. and Kempka, R. (2024) "Towards Advancing Real-Time Railroad Inspection Using a Directional Eddy Current Probe".
- Office of Rail and Road (2024) *Annual efficiency and finance assessment of Network Rail 2024*. Doi: <https://www.orr.gov.uk/annual-efficiency-and-finance-assessment-network-rail-2024>.
- Office of Railroad Safety (2015) *Track Inspector Defect Reference Manual*, Federal Railroad Administration.
- Ou, W., et al., 2025. A real-time image retrieval and localization method based on 360-degree panoramic visual feature maps. *ISPRS J. Photogram. Remote Sens.* 229, 351–365. <https://doi.org/10.1016/j.isprsjprs.2025.08.018>.
- Pham, D., Ha, M., Xiao, C., 2021. A novel visual inspection system for rail surface spalling detection. *IOP Conf. Series: Mater. Sci. Eng.* 1048 (1). <https://doi.org/10.1088/1757-899X/1048/1/012015>.
- Porsch, R., Lehmann, M.A., 2023. *Evaluation of marker-based AR-tracking with vuforia in the context of rail vehicle maintenance*. In: *Lecture Notes in Networks and Systems*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28076-4_7.
- Portalés, C., Lerma, J.L., Navarro, S., 2010. Augmented reality and photogrammetry: A synergy to visualize physical and virtual city environments. *ISPRS J. Photogram. Remote Sens.* 65 (1), 134–142. <https://doi.org/10.1016/j.isprsjprs.2009.10.001>.
- Qin, C., et al., 2022. Concentrated velocity synchronous linear chirplet transform with application to robotic drilling chatter monitoring. *Measurement* 194. <https://doi.org/10.1016/j.measurement.2022.111090>.
- Rahman, M., et al., 2023. A railway track reconstruction method using robotic vision on a mobile manipulator: A proposed strategy. *Comp. Indust.* 148, 103900. <https://doi.org/10.1016/j.compind.2023.103900>.
- Restrepo-Girón, A.D., Loaiza-Correa, H., 2016. Background thermal compensation by filtering for contrast enhancement in active thermography. *J. Nondestruct. Eval.* 35 (1), 20. <https://doi.org/10.1007/s10921-016-0336-x>.
- Rothenberger, N., et al., 2025. Illumination invariant image matching for lunar TRN. In: *AIAA SCITECH 2025 Forum*. Reston, Virginia: American Institute of Aeronautics and Astronautics. <https://doi.org/10.2514/6.2025-2073>.
- Scheffer, S., et al., 2021. Augmented reality for IT/OT failures in maintenance operations of digitized trains: current status, research challenges and future directions. *Procedia CIRP* 100, 816–821. <https://doi.org/10.1016/j.procir.2021.05.038>.
- Sellán, S., Jacobson, A., 2022. Stochastic poisson surface reconstruction. *ACM Trans. Grap.* 41 (6), 1–12. <https://doi.org/10.1145/3550454.3555441>.
- Shapiro, R., 1978. Direct linear transformation method for three-dimensional cinematography. *Res. Quart. Am. Alliance Health, Phys. Edu. Recreation* 49 (2), 197–205. <https://doi.org/10.1080/10671315.1978.10615524>.
- Shen, Y., et al., 2022. GCN-denoiser: mesh denoising with graph convolutional networks. *ACM Trans. Grap.* 41 (1), 1–14. <https://doi.org/10.1145/3480168>.
- Sun, L., et al., 2022. In: *Event-Based Fusion for Motion Deblurring with Cross-modal Attention*, pp. 412–428. https://doi.org/10.1007/978-3-031-19797-0_24.
- Tareen, S.A.K., Saleem, Z., 2018. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, pp. 1–10. <https://doi.org/10.1109/ICOMET.2018.8346440>.
- VOSSLOH (2025) *SoniQ Rail Explorer: Ultrasonic Inspection System*. Doi: <https://vossloh-topic-lounge.com/smart-maintenance-track-supply/soniq-rail-explorer/>.
- Wang, J., et al., 2023. A fast, efficient, and explicit phase-field model for 3D mesh denoising. *Appl. Math. Comput.* 458. <https://doi.org/10.1016/j.amc.2023.128239>.
- Wang, R., et al., 2024a. Real-time UWB and IMU fusion positioning system for urban rail transit with high mobility. *IEEE Veh. Technol. Conf.* 1–7. <https://doi.org/10.1109/VTC2024-Spring62846.2024.10683298>.
- Wang, Y., et al., 2024b. A full 3D reconstruction of rail tracks using a camera array. *Measurement* 225. <https://doi.org/10.1016/j.measurement.2023.114034>.
- Wang, Y., et al., 2025. Multiple-modality sensor fusion for rail track 3D reconstruction. *Computers in Industry*.
- Wang, Y., et al., 2026. Research on geometric parameter quantification of rail rolling contact fatigue crack damage based on 2D optical image. *Scient. Rep.* <https://doi.org/10.1038/s41598-026-36276-w>.
- Wei, Y., et al., 2019. A non-contact measurement method of ship block using image-based 3D reconstruction technology. *Ocean Eng.* 178, 463–475. <https://doi.org/10.1016/j.oceaneng.2019.03.015>.
- Winsor, T., Bourn, J. and Hull, G. (2000) *Public Accounts - Thirty-Fifth Report*. Doi: <https://publications.parliament.uk/pa/cm199900/cmselect/cmpubacc/536/53603.htm>.
- Wu, C., 2013. Towards linear-time incremental structure from motion. In: *Proceedings - 2013 International Conference on 3D Vision, 3DV 2013*, pp. 127–134. <https://doi.org/10.1109/3DV.2013.25>.
- Wu, C., et al., 2021. Which accuracy levels of positioning technologies do drivers really need in connected vehicle settings for safety? *Accident Analysis and Prevention* 157. <https://doi.org/10.1016/j.aap.2021.106106>.
- Xiong, L., et al., 2023. Detection of rail defects using NDT methods. *Sensors* 23 (10), 4627. <https://doi.org/10.3390/s23104627>.
- Xu, J., Doyle, D. and Moreo, F. (2023) "State of the art of augmented reality capabilities for civil infrastructure applications," (October 2022), pp. 1–16. Doi: 10.1002/eng2.12602.
- Xu, X., et al., 2023. Object detection algorithm for railway scenes based on infrared and RGB image fusion. In: *2023 International Conference on Pattern Recognition, Machine Vision and Intelligent Algorithms (PRMVIA)*. IEEE, pp. 52–58. <https://doi.org/10.1109/PRMVIA58252.2023.00015>.
- Xue, Y., et al., 2021. Novel SfM-DLT method for metro tunnel 3D reconstruction and Visualization. *Underground Space* 6 (2), 134–141. <https://doi.org/10.1016/j.undsp.2020.01.002>.
- Xue, Z., et al., 2023. Systematic review: Ultrasonic technology for detecting rail defects. *Const. Build. Mater.* 368. <https://doi.org/10.1016/j.conbuildmat.2023.130409>.
- Yang, X., et al., 2024. Research on experimental method of high-speed railway perimeter intrusion monitoring technology based on multi-sensor fusion. In: *Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management*. ACM, New York, NY, USA, pp. 206–214. <https://doi.org/10.1145/3718751.3718784>.
- Yaopu, Z., Yanchao, Y., Ying, L., 2018. Research on Construction Technology of 3D Scene Point Cloud Model Based on Improved PMVS Algorithm. In: *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*. ACM, New York, NY, USA, pp. 359–363. <https://doi.org/10.1145/3297156.3297192>.
- Zhang, D., Lingamanaik, Siva N., Chung, H., 2023a. Image-based 3D reconstruction for rail profile measurement. *Proc. Inst. Mech. Eng. Part F: J. Rail Rapid Transit* 237 (3), 309–321. <https://doi.org/10.1177/09544097221110322>.
- Zhang, D., Lingamanaik, Siva N., Chung, H., 2023b. Image-based 3D reconstruction for rail profile measurement. *Proc. Inst. Mech. Eng. Part F: J. Rail Rapid Transit* 237 (3), 309–321. <https://doi.org/10.1177/09544097221110322>.
- Zhang, H., et al., 2024. Characterization of multi-layer rolling contact fatigue defects in railway rails using sweeping eddy current pulse thermal-tomography. *Appl. Sci.* 14 (16), 7269. <https://doi.org/10.3390/app14167269>.
- Zhang, X., et al., 2013. Edge strength similarity for image quality assessment. *IEEE Signal Proc. Lett.* 20 (4), 319–322. <https://doi.org/10.1109/LSP.2013.2244081>.
- Zhang, Z. et al. (2022) "An Empirical analysis of freight train derailment rates for unit trains and manifest trains," 236(10), pp. 1168–1178. Doi: 10.1177/09544097221080615.
- Zhou, D., et al., 2023. Feasibility study of deep learning-based markerless real-time lung tumor tracking with orthogonal X-ray projection images. *J. Appl. Clin. Med. Phys.* 24 (4). <https://doi.org/10.1002/acm2.13894>.
- Zhou, Y., et al., 2025. Estimating railway track curvature using gyroscope and GPS sensors. *Scient. Rep.* 15 (1), 1–12. <https://doi.org/10.1038/s41598-025-91255-x>.



DSPACE

<https://dspace.org/>

Railway track surface defect 3D reconstruction and relocalisation through cross-modality fusion

Wang, Yizhong; Yang, Lichao; Durazo-Cardenas, Isidro; Zhao, Yifan

2026-05

Elsevier

<https://doi.org/10.1016/j.isprsjprs.2026.03.021>

Wang Y, Yang L, Durazo-Cardenas I, Zhao Y. (2026) Railway track surface defect 3D reconstruction and relocalisation through cross-modality fusion. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 235, May 2026, pp. 618-633