applied sciences

MDPI

Article

Development and Experimental Testing of a 3D Vision System for Railway Freight Wagon Monitoring

Alessio Cascino [1], Simone Delle Monache [2], Laurens Lanzillo [3], Francesco Mazzeo [3], Leandro Nencioni [1], Armando Nicolella [4], Salvatore Strano [4,*] and Mario Terzo [4]

[1] Department of Industrial Engineering, University of Florence, 50139 Florence, Italy; alessio.cascino@unifi.it (A.C.); leandro.nencioni@unifi.it (L.N.)
[2] Mercitalia Intermodal SpA, 20151 Milano, Italy; s.dellemonache@mercitaliaintermodal.it
[3] Department of Mechanical Engineering, Politecnico di Milano, 20156 Milan, Italy; laurens.lanzillo@polimi.it (L.L.); francesco.mazzeo@polimi.it (F.M.)
[4] Department of Industrial Engineering, University of Naples, 80125 Naples, Italy; armando.nicolella@unina.it (A.N.); mario.terzo@unina.it (M.T.)
* Correspondence: salvatore.strano@unina.it

Abstract

Ensuring the safety and reliability of freight wagons requires continuous monitoring of couplings such as hooks and buffers, which are prone to stress, wear, and misalignments. This paper proposes a vision-based 3D monitoring system which uses an RGB-D camera and a computer vision pipeline to estimate angular excursions and longitudinal displacements of wagon couplers during train operation. The proposed approach combines depth-based reconstruction with a normalized cross-correlation tracking algorithm, providing geometric measurements of coupling motion without physical contact. The system architecture integrates real-time acquisition and post-processing analysis to 3D reconstruct the geometric characteristics of wagon couplings under field conditions. Experimental tests performed on a T3000 articulated wagon allowed us to measure an angular excursion of approximately 9.8° for the hook and a longitudinal displacement of 17 mm for the buffer. The results show robustness and suitability for embedded implementation, supporting the adoption of vision-based techniques for safety monitoring in railways.

Keywords: computer vision; RGB-D sensing; structural health monitoring; railway freight wagons; coupler (hook–buffer) monitoring; template matching-based tracking

check for updates

Academic Editor: Douglas O'Shaughnessy

Received: 1 October 2025
Revised: 20 October 2025
Accepted: 22 October 2025
Published: 29 October 2025

Citation: Cascino, A.; Delle Monache, S.; Lanzillo, L.; Mazzeo, F.; Nencioni, L.; Nicolella, A.; Strano, S.; Terzo, M. Development and Experimental Testing of a 3D Vision System for Railway Freight Wagon Monitoring. *Appl. Sci.* **2025**, *15*, 11547. https://doi.org/10.3390/app152111547

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rail freight transport is a key part of modern logistics, offering a sustainable, high-capacity, and cost-effective alternative to road transport. In line with the European Union's strategic objectives of decarbonization and decongestion, recent policies have encouraged the use of trains to increase network capacity [1,2]. In this context, the ability to ensure operational safety and structural reliability of freight wagons takes relevance.

Freight rail operations rely on mechanical couplings (such as buffers, hooks, and drawgears) to keep the physical link between wagons and to ensure the proper dynamic of the entire freight vehicle [3,4]. These components are responsible for both load transmission and dynamic stability and are of great importance; indeed, any malfunction, deformation, or misalignment in these components can compromise the safety of the operations and create economic consequences across the logistics chain. A broken coupling or knuckle can cause a train to separate, triggering an emergency brake and

potentially resulting in derailments [5–7]. Such events are dangerous (especially if they involve hazardous cargo) and lead to operational disruption. For example, in heavy-haul trains, the lateral stability of couplers is closely monitored because even slight angular deviations can threaten train handling and stability [8]. This suggests that one of the most important things is to monitor coupling properly, from both a safety and an operational point of view.

Traditionally, the condition of mechanical couplings and related wagon components has been assessed through periodic manual inspections and onboard sensing devices. Maintenance staff typically perform visual checks on couplers, looking for signs of wear, cracks, or incorrect engagement during pre-departure inspections. However, manual inspections are intensive, subjective, and limited in frequency, which means that issues can occur and go unnoticed between scheduled checks. Visibility is often constrained (e.g., poor lighting or inaccessible inspection angles), and inspectors might miss defects until they escalate.

Driven by these limitations, the rail industry is increasingly turning toward real-time monitoring, condition-based maintenance strategies, and railways damage detection [9,10]. Instead of fixed inspection intervals, condition-based maintenance calls for servicing equipment when data indicates the need [11]. This approach can be really pertinent for freight couplings and suspension elements [12,13], which experience highly variable stress cycles. Real-time monitoring systems can track coupling performance (e.g., forces, displacements, and alignment) during train operation and alert operators to anomalies [14–16]. Recent technological advances have made such continuous monitoring feasible; for example, digital retrofitting of freight wagons with sensors and wireless communication units allows live data on wagon health to be collected and analyzed. Intelligent freight wagons have demonstrated the value of this data-driven approach, detecting faults and wear conditions that would not have been caught during periodic checks [17].

Among emerging monitoring technologies, computer vision has gained attention as a non-invasive and cost-effective solution to improve safety in the normal functioning of trains [18–20]. One of the main targets of the onboard vision systems can be for assessing the condition of mechanical components. Vision-based monitoring uses cameras and image analysis algorithms to visually inspect components, which in the context of freight wagons means monitoring the couplers, buffers, and other parts for signs of damage or abnormal behavior.

A key advantage of using cameras is that physical contact is not required to monitor the components. This means that the sensor does not interfere with the normal functioning of the mechanical system and avoids issues of sensor durability. Instead, cameras can be positioned either on board or on the wayside to capture images or video of the wagon connections in service. Vision systems can directly measure the relative motion or alignment between mechanical components [21,22]. Tracking visual markers or features on consecutive wagons can be useful to quantify displacements or angular rotations at the coupling; with the ongoing digitalization of rail assets and improvements in image processing, such vision techniques are becoming increasingly practical for railway maintenance [23]. Computer vision offers a complementary approach to traditional sensors, providing continuous visual data about coupling conditions without direct sensor installation on every component [24].

In recent years, computer vision has therefore emerged as one of the main technologies driving automation in the railway sector. Early approaches, such as that by [25], demonstrated the feasibility of deep learning use for rail track monitoring, paving the way for the new AI-based generation. Since then, several solutions have used Convolutional Neural Networks (CNNs) and multimodal sensing to improve detection accuracy and

environmental robustness. In [26,27], CNN-based systems for rail surface and defect detection have been developed, achieving high precision under controlled conditions. In [27], a model has been proposed to reduce computational complexity, enabling a near real-time inference on embedded GPUs. In [28], it is possible to find a comprehnsive study on the use of CNNs for track maintenance, highlighting the need for a reliable generalization under variable lighting conditions. More recent developments have focused on enhancing computational efficiency and enabling real-time onboard operation. The method proposed in [29] is a lightweight convolutional network for missing bolt detection, optimized for deployment on resource-limited embedded devices. The study by [30] presents RH-Net, a CNN–transformer hybrid architecture that achieved high accuracy and real-time performance for obstacle detection. The study by [31] improved generalization through the use of multiple CNNs across multiple track conditions, while [32] proposed an FPGA-based AI solution for fault detection that balances energy efficiency and computational speed. Hybrid vision–sensor systems have also been explored: The authors of [33] developed a wayside AI-assisted vision system that integrates optical inspection with vibration sensing. The authors of [34] implemented an unsupervised anomaly detection framework based on autoencoders, identifying visual irregularities without prior labeling. Further lightweight approaches, such as that by [35], further demonstrate the trend toward embedded, real-time, and efficient visual inspection systems. Despite these advances, many limitations remain: many current systems depend on high-end computing or offline data processing, which affects their scalability for deployment. Moreover variations in illumination, vibration, and occlusion remain key challenges affecting robustness. While significant progress has been achieved, there is still a clear need for cost-effective, onboard, and real-time vision systems capable of maintaining reliable performance under operating conditions. Based on these considerations, this study proposes a vision-based monitoring framework designed to overcome limitations such as dependence on high-end hardware, limited adaptability, and a lack of 3D motion estimation. This work introduces a practical computer vision approach designed for onboard implementation.

The main objective of this study is to develop a vision-based 3D monitoring system capable of estimating the angular excursions and longitudinal displacements of mechanical couplings in freight wagons during operation in order to support condition-based maintenance and enhance safety under real field conditions. Therefore, the main contributions can be summarized as follows:

- Development of an RGB-D vision-based system for 3D monitoring of railway wagon couplings, designed for real operating conditions.
- Integration of depth-based reconstruction with a normalized cross-correlation tracking algorithm to compute 3D geometric measurements without physical contact.
- Experimental validation by deploying the proposed methodology on a T3000 articulated freight wagon, thus demonstrating measurements consistent with operational standards.
- Design of a processing pipeline suitable for embedded implementation, supporting condition-based maintenance and safety monitoring strategies in rail transport.

This paper is organized as follows: In Section 2, the system setup is illustrated; the computer vision algorithm is described in Section 3; experimental results are presented in Section 4. Finally, conclusions and future developments are drawn in Section 5.

2. System Setup

The experimental setup is installed on the extremity of a T3000 articulated wagon. The T3000 wagon is a modern articulated intermodal freight wagon manufactured by *Ferriere Cattaneo SA* (Gubiasco, Switzerland) for the transport of both semi-trailers and

containers. Figure 1 details the position of the setup and highlights the buffer and hook which are seen by the camera.

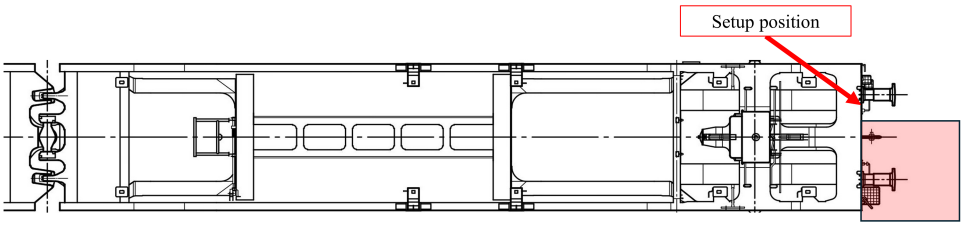


Figure 1. Setup position on the wagon.

The setup involves an RGB-D vision system designed for monitoring railway freight wagons, buffers, and hooks (Figure 2) mounted on the wagon chassis and oriented toward the coupler region. The camera is connected to an HP compact computing unit (NUC type) running Ubuntu and ROS2, which manages the real-time acquisition and synchronization of data streams.

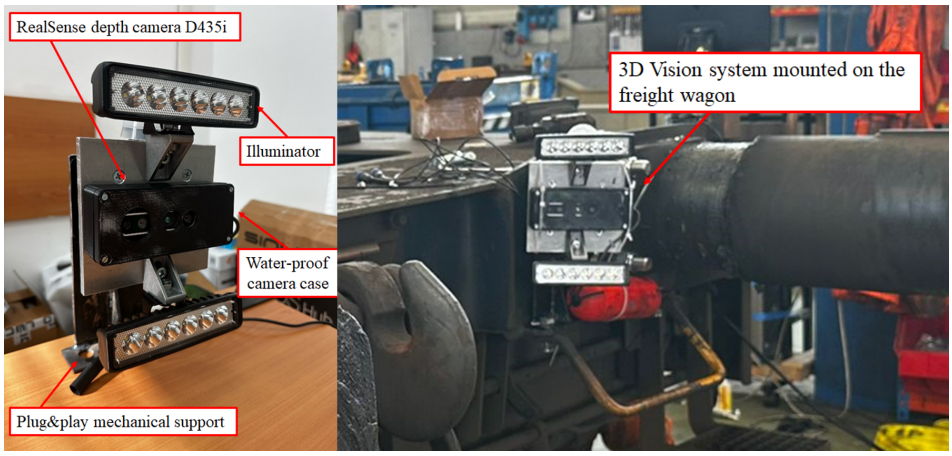


Figure 2. System setup.

The camera is configured and installed upside down for operational requirements, involving a 180-degree rotation of captured frames in subsequent data processing stages. The post-process stage begins with a Python script developed using the pyrealsense2 library (Intel RealSense SDK 2.0, Intel Corporation, Santa Clara, CA, USA; accessed on 10 July 2024). script developed to handle data recorded in a compressed .bag file format. This script extracts synchronized RGB and depth images, aligns them spatially and temporally, and saves these images as structured MATLAB data (MATLAB R2024a, MathWorks, Natick, MA, USA). (.mat files) for further processing. In our setup, the RealSense D435i camera (in Figure 3) recorded synchronized RGB (640 × 480) and depth (640 × 480) frames at 15 fps, ensuring uniform temporal sampling for all the following analyses.

Intrinsic parameters for 3D reconstruction, such as focal lengths (fx, fy) and principal points (cx, cy), are also extracted from the calibration process and stored. In matrix form, these scalars are compacted into the camera matrix K, where

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

In addition to the intrinsic parameters, the calibration process also provided the extrinsic parameters and the lens distortion coefficients. The extrinsic parameters define the transformation matrix that relates the right stereo camera to the left one, while the

distortion coefficients characterize the radial and tangential distortions introduced by the lens. These parameters are essential for accurately back-projecting the 2D tracked points into 3D space, ensuring a geometric analysis. Environmental factors influenced the setup significantly: a transparent protective panel placed in front of the camera to achieve protection standards introduced potential image distortions; therefore, post-process steps were implemented to reduce this effect. In order to ensure measurement accuracy, the calibration process was carried out with the protective housing already installed in front of the camera using a standard chessboard calibration pattern. This procedure compensated for any optical distortions introduced by the transparent panel and guaranteed the correctness of the camera parameters. The camera and supporting structure were rigidly mounted to minimize vibration effects and mechanical oscillations during train motion. The experiments were conducted under natural daylight conditions, with moderate variations in illumination typical of outdoor rail environments. Despite minor motion blur and lighting changes, the combination of frame rate and depth-based reconstruction ensured stable feature tracking performance throughout the test.

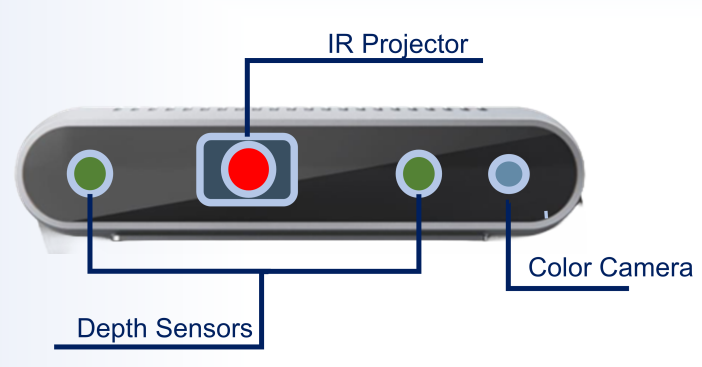


Figure 3. The Realsense D435i camera used in the test, highlighting the camera sensors used for the algorithm.

The functioning architecture (Figure 4) of the system can be divided into three main modules: (1) real-time image acquisition, data recording and timestamping, (2) offline feature extraction and motion estimation, and (3) post-processing and visualization of results.

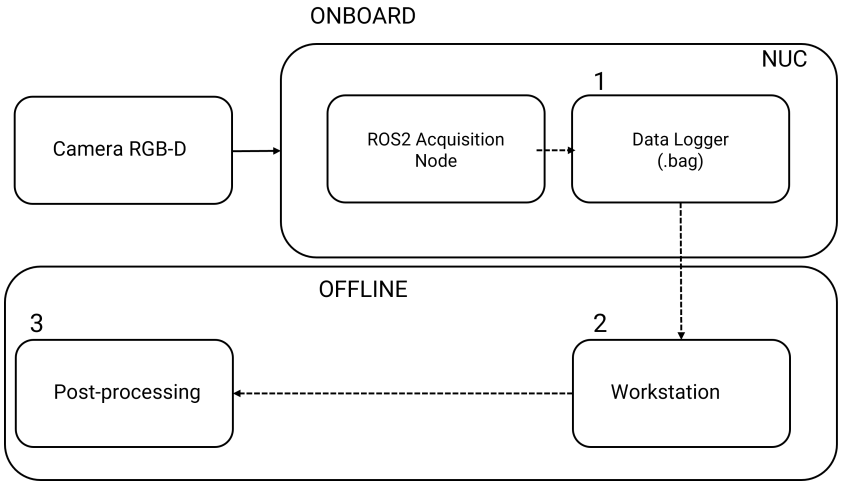


Figure 4. System architecture.

The monitored scene consists of specific mechanical components of railway freight wagons, particularly hooks and buffers, chosen for better analysis of the relative movements between two carriages and for their relevance to the structural integrity connection and

safety of railway operations. The positioning of the camera was strategically selected to provide better coverage of these components, ensuring the analysis of their displacement and orientation during motion. Moreover, such a choice minimized interferences with normal operations performed on the wagon, such as coupling and uncoupling with adjacent wagons or loading operations.

3. Description of the Computer Vision Method

The computer vision algorithm developed for this system is structured as a multistage pipeline that integrates manual picking input with automated detection and tracking, as shown in Figure 5.

The main processing steps of the proposed algorithm, as illustrated in Figure 5, are summarized as follows:

1. Acquire synchronized RGB and depth frames.
2. Manually select points of interest on the initial RGB frame.
3. For each frame k:
   - 3.1. Perform normalized cross-correlation (NCC) within a region of interest to track template 2D positions.
   - 3.2. Extract valid depth values and back-project the tracked points to 3D coordinates.
   - 3.3. Compute the hook angle $\theta(k)$ and the buffer displacement $p_{mid}(k)$.
4. Apply a moving-average filter to smooth signals and compute $\Delta\theta$ and Δp.

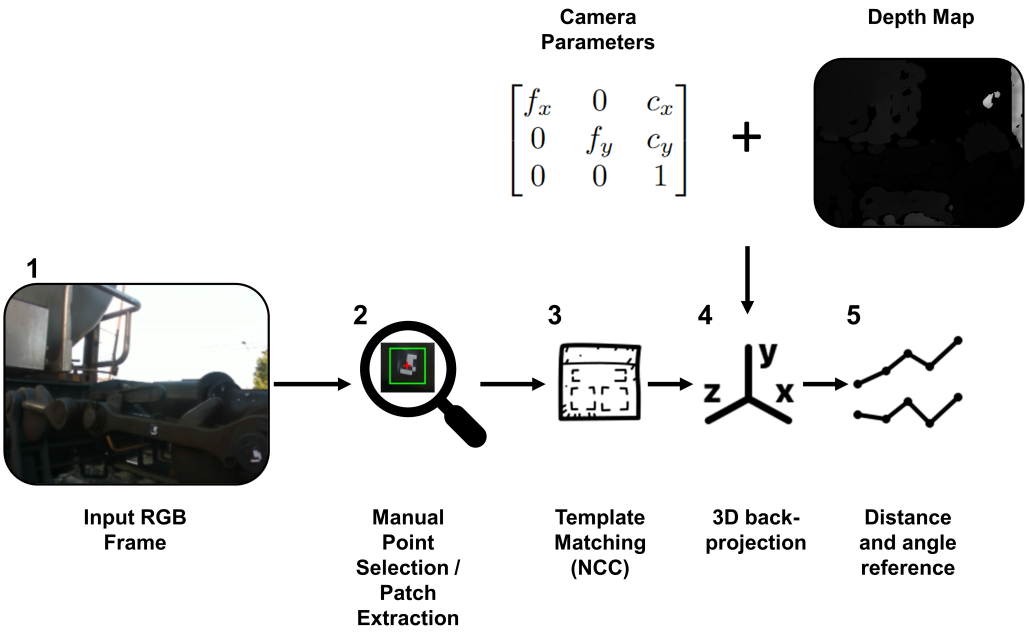


Figure 5. Overview of the developed computer vision pipeline.

For every new frame, tracking is performed using a template matching approach based on normalized cross-correlation (NCC) [36]. For each tracked point, the algorithm searches for the best match of the reference path within a region of interest (ROI), in the current frame, centered around the last known 2D position. The algorithm computes the NCC between the template and candidate regions, selecting the location with the highest score as the updated position of the point:

$$(u^*, v^*) = \underset{(u,v)}{\operatorname{argmax}} \left(\frac{\sum_{x,y}(T(x,y) - \bar{T})(I(x+u, y+v) - \bar{I}_{u,v})}{\sqrt{\sum_{x,y}(T(x,y) - \bar{T})^2 \sum_{x,y}(I(x+u, y+v) - \bar{I}_{u,v})^2}} \right) \quad (2)$$

where (u^*, v^*) represents the coordinates of the point that is recognized with the maximum correlation; $T(x, y)$ is the template patch; $I(x + u, y + v)$ is the ROI in the current frame; and $\bar{T}$ and $\bar{I}_{u,v}$ represent their respective mean intensities. This approach, while simple, is known for its effectiveness in relatively static or rigid scenes.

After each point is located in the 2D image, its depth value is extracted from the corresponding depth map. If the depth at that exact pixel is zero or otherwise invalid, the algorithm falls back on a local search: it scans a square neighborhood of size $(w \times w)$ centered on the pixel ($w = 31$ px default pixels, i.e., a radius of 15 px). The valid depth sample whose coordinates are closest to the center of this window is selected. Afterwards, each tracked 2D point is back-projected into 3D space using the previous depth information from the aligned depth images and the obtained intrinsic camera parameters:

$$\begin{aligned} X &= \frac{(u^* - c_x)}{f_x} \cdot z \\ Y &= \frac{(v^* - c_y)}{f_y} \cdot z \\ Z &= z \end{aligned} \tag{3}$$

where z is the depth information. Using 3D information, it is possible to measure geometric changes and structural movement. Once two points are tracked (each point selected belongs to the hook in order to define an axis parallel to the axis of the hook), the algorithm can track the angle that the line connecting them forms with the optical axis (Z-axis) of the camera (hook case) and the projection along the object axis of the midpoint (buffer case). Given the 3D coordinates of the two tracked points at frame k, $\mathbf{p}_1^{(k)}$ and $\mathbf{p}_2^{(k)}$, the direction vector is computed, and the normalized direction is

$$\begin{aligned} \mathbf{v}^{(k)} &= \mathbf{p}_2^{(k)} - \mathbf{p}_1^{(k)} \\ \hat{\mathbf{v}}^{(k)} &= \frac{\mathbf{v}^{(k)}}{||\mathbf{v}^{(k)}||} \end{aligned} \tag{4}$$

The angle $\theta^{(k)}$ between the segment and the camera Z-axis $\hat{\mathbf{z}}$ is then calculated as

$$\begin{aligned} \cos\theta^{(k)} &= \hat{\mathbf{v}}^{(k)} \cdot \hat{\mathbf{z}} \\ \theta^{(k)} &= \arccos\left(\hat{\mathbf{v}}^{(k)} \cdot \hat{\mathbf{z}}\right) \end{aligned} \tag{5}$$

The angle in consideration is shown in Figure 6.

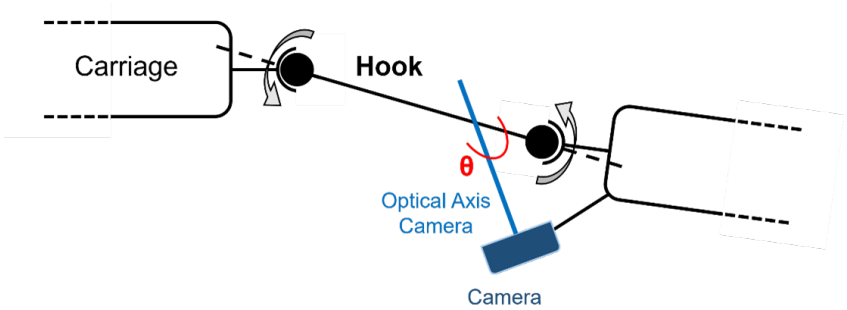


Figure 6. Top view of the hook analysis.

While this angle can be transformed into the carriage reference frame if needed, for the purpose of this study, the primary scope is the angular excursion, that is, the range of angle values observed throughout the sequence.

$$\Delta\theta = \theta_{\max} - \theta_{\min} \tag{6}$$

This metric quantifies the rotation experienced by the monitored component, without having an explicit transformation to the wagon reference system. As a result, the maximum angular excursion is used as the practical index for monitoring changes in the geometry of the monitored system. On the other hand, Figure 7 shows the displacement d for the buffer along its main axis.

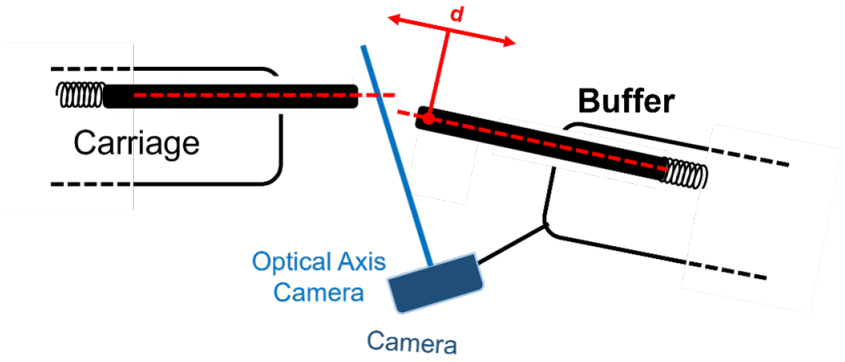


Figure 7. Top view of the buffer analysis.

For each frame k, the algorithm begins with the computation of the midpoint between the two tracked points (as conducted before, the two selected points belong to the buffer in order to draw an axis parallel to the buffer axis):

$$\mathbf{m}^{(k)} = \frac{1}{2}\left(\mathbf{c}_1^{(k)} + \mathbf{c}_2^{(k)}\right) \tag{7}$$

where $\mathbf{c}_1^{(k)}$ and $\mathbf{c}_2^{(k)}$ denote the 3D coordinates of the two points at frame k. The axis of interest is defined by the line connecting the mean positions of the two point clouds across all frames, where

$$\bar{\mathbf{c}}_1 = \frac{1}{N}\sum_{k=1}^{N}\mathbf{c}_1^{(k)}, \quad \bar{\mathbf{c}}_2 = \frac{1}{N}\sum_{k=1}^{N}\mathbf{c}_2^{(k)} \tag{8}$$

$$\mathbf{d} = \frac{\bar{\mathbf{c}}_2 - \bar{\mathbf{c}}_1}{\|\bar{\mathbf{c}}_2 - \bar{\mathbf{c}}_1\|} \tag{9}$$

The projection of the midpoint at frame k onto the reference axis with respect to $\bar{\mathbf{c}}_1$ is then calculated as

$$p_{\text{mid}}^{(k)} = \left(\mathbf{m}^{(k)} - \bar{\mathbf{c}}_1\right) \cdot \mathbf{d} \tag{10}$$

As performed for the angle, the excursion of the displacement is then computed as

$$\Delta p = p_{\max} - p_{\min}, \quad \text{where} \quad p_{\max} = \max_k p_{\text{mid}}^{(k)}, \quad p_{\min} = \min_k p_{\text{mid}}^{(k)} \tag{11}$$

The pipeline is designed to be adaptable and easily accessible. It can be extended or integrated with other sensing modalities or high-level decision-making frameworks. The algorithm shows how combining RGB-D data, system calibration, and computer vision techniques can be implemented for practical and effective solutions for 3D structural monitoring in challenging, real-world environments such as railway operations.

4. Experimental Results

The experimental results presented in this section show the effectiveness of the proposed 3D tracking and analysis pipeline in a railway context. Scenario 1 focuses on the angular excursion of the hook axis with respect to the camera optical axis, while Scenario 2 analyzes the longitudinal displacements of the buffer along its main axis. Therefore, the whole analysis concentrates on the measurement of these angular and longitudinal variations recorded during operations. Such variations are interpreted as the actual dynamic response of the hook and buffer under real field conditions, rather than as deviations from predefined nominal parameters. These two cases represent key quantities for assessing coupling integrity and mechanical stability during wagon operation.

In order to gather the necessary data, the instrumented wagon was inserted in a freight trainset running on the Milano–Bari line in Italy. The trainset was composed of a single locomotive and 19 wagons, with the instrumented one in the middle of the composition. The post-process analysis was made considering an acquisition of approximately 33 s in the line section highlighted in Figure 8. This line section is mainly straight, and it is traveled by wagons at an average speed of 100 km/h. A dedicated sensor installed at the braking cylinder test point of the wagon confirmed that the train was not braking during the maneuver.

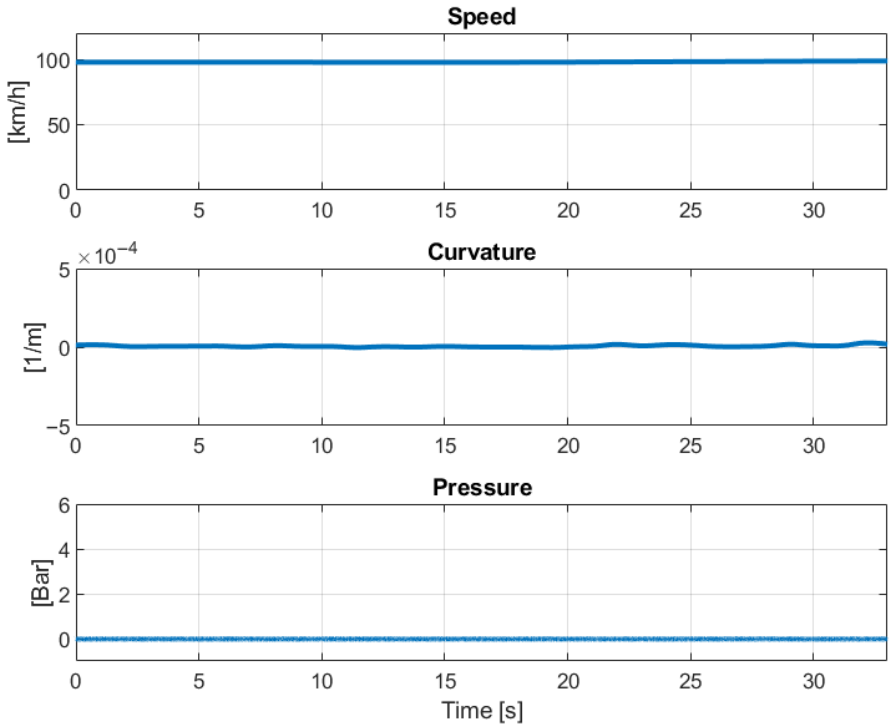


Figure 8. Trainset speed, line curvature, and brake cylinder pressure in the section considered during the acquisition phase.

As mentioned, for the hook (Figure 6), the focus was on monitoring possible rotations or misalignments, which were quantified through the analysis of angular excursion. In contrast, for the buffer, the main interest was in tracking longitudinal displacement along the object axis (Figure 7). In both cases, the raw measurements were post-processed using a moving average filter to suppress noise and highlight physically meaningful trends. The general form of the moving average for a discrete sequence $x[n]$ is given by

$$\tilde{x}[n] = \frac{1}{N} \sum_{k=-M}^{M} x[n+k] \tag{12}$$

where $N = 2M + 1$ is the window size and M is the half-window. In this specific case, few considerations have been assumed: given a train speed of 70 km/h and a camera frame rate of 15 fps, the distance covered in each frame is approximately 1.3 m. Considering that the standard length of a rail module is at least 12 m, the train covers 9.1 m in seven frames. Therefore, using a window of seven frames ensures that the moving average is computed over a segment shorter than the length of a single module. This choice provides a good compromise between signal smoothing and temporal resolution, as it reduces noise while preserving relevant changes related to each rail module.

4.1. Case: The Angle Between the Camera Optical Z-Axis and the Hook Axis

The first experimental scenario deals with the monitoring of rotation behavior of the hook, that is, the rigid mechanical component coupling two railway carriages. Figure 9 shows the projection of the measured 3D axis and tracked points onto the camera image.



Figure 9. Projection of the 3D hook axis (in red) and a pair of tracked points (blue and green markers) on the RGB image.

For this case, the critical metric is the angle θ formed between the 3D line connecting the two tracked points on the hook and the optical Z-axis of the camera, as shown in Figure 6. This measurement provides insight into potential rotations during the normal running of the train and becomes a parameter of primary interest for structural and safety monitoring. As the train moves, the angle θ was computed for each frame, and its evolution is plotted in Figure 10.

The result of this analysis is the excursion $\Delta\theta$, defined as the difference between the maximum angle θ_{max} and the minimum angle θ_{min} measured during the sequence.

The filtered signal shows an excursion of 9.8°: the excursion can be interpreted as the actual range of rotational motion experienced by the hook during the monitoring period. In practical terms, this value can suggest misalignment or relative hunting motion between the two carriages and can be used as real-time alert systems in operational railway contexts. The measured filtered angular excursion of 9.8° falls within the range of operational values documented in the literature for freight wagon couplings under operational conditions [37].

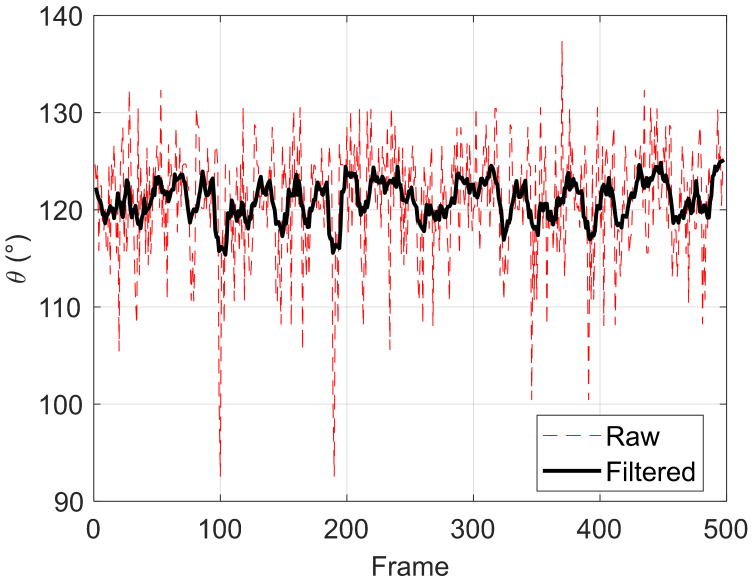


Figure 10. Angle between the 3D hook axis and the optical camera axis.

4.2. Case: Displacement Along the Buffer Axis

The second experimental scenario focuses on monitoring longitudinal displacement in the buffer, a structural component designed to absorb and compensate for relative motion between rail carriages. In this case, the metric is the displacement of the midpoint along the buffer axis. Figure 11 shows the projection of the 3D object axis and the tracked points onto the camera image, providing a visual representation of the measured geometry. At each frame, the instantaneous midpoint between the tracked points is computed, and its projection onto the mean axis is taken: this operation yields the displacement signal $p_{\text{mid}}^{(k)}$, which characterizes the dynamic response of the buffer.



Figure 11. Projection of the 3D buffer axis (in red) and a pair of tracked points (blue and green markers) on the RGB image.

The evolution of this displacement is shown in Figure 12, where both raw and filtered signals are plotted. The unfiltered displacement has high-frequency oscillations, reflecting both measurement noise and small tracking errors inherent to real-world conditions.

Application of the moving average filter reduces this variability, allowing the detection of trends and shifts which maybe otherwise be masked.

The range of the filtered displacement signal is about 17 mm. This displacement is within the European standards, which state that freight wagon buffers typically have stroke capacities of 105–110 mm for standard applications. The displacement excursion provides an index for assessing buffer performance and identifying possible abnormal events.

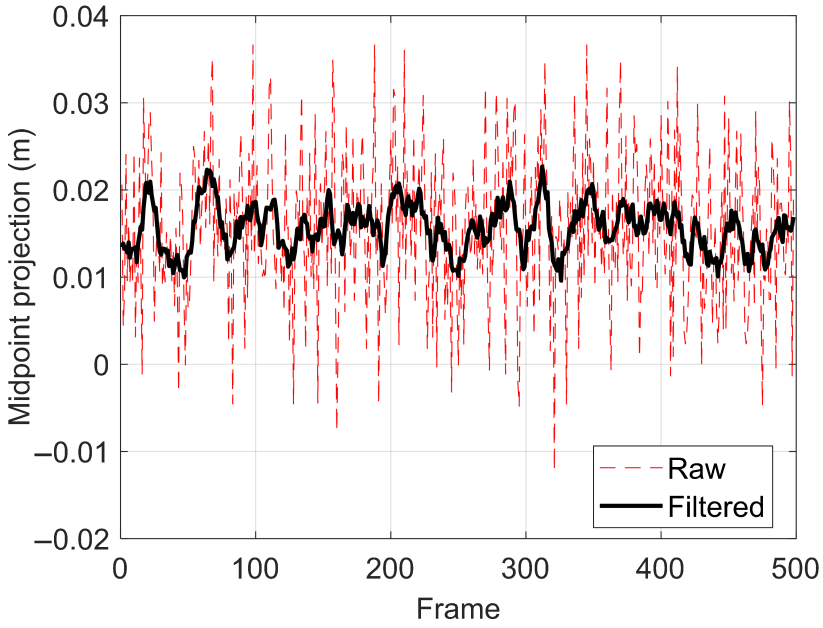


Figure 12. Midpoint displacement along the buffer axis.

5. Discussion

The experimental results confirm the capability of the proposed vision-based system to measure the angular excursion and longitudinal displacements of the freight wagon couplings under real operating conditions. Compared with other computer vision approaches for railway inspection [26,27,29,30], the proposed system focuses specifically on the 3D monitoring of mechanical connections, which has received limited attention in the literature. Unlike existing image-based defect detection systems that rely on 2D analysis or require defined environments, the proposed method integrates RGB-D sensing to reconstruct real geometric quantities in dynamic field conditions. When compared to lightweight CNN-based systems such as those proposed in [29,30], the approach proposed in this paper does not require a learning phase, avoiding dependency on large datasets. This makes it more adaptable to new environments and mechanical configurations. Nevertheless, some limitations must be considered. Indeed, the accuracy of the 3D reconstruction can be affected by illumination changes and partial occlusions caused by vibrations or dust. However, in the proposed setup, the regions of interest were selected over feature-rich areas, ensuring robust correlation even under variable lighting. This strategy mitigated the typical sensitivity of the NCC algorithm to brightness fluctuations, occlusions, and scale changes, resulting in consistent tracking performance throughout the field tests and better point tracking over time. The depth accuracy of the Intel RealSense D435i camera was also considered. The camera was recalibrated together with its protective housing, and the corrected intrinsic parameters were used for depth map computation. The built-in infrared (IR) projector generates a structured-light pattern that enhances depth estimation reliability in low-texture regions. Although depth uncertainty can propagate through the 3D reconstruction process, its impact on the angular measurement θ is limited due to normalization.

The longitudinal displacement measurement may be slightly more affected; however, given the tested range and the camera's specified accuracy (less than 2% at 2 m), the resulting 3D estimates are sufficiently precise for monitoring purposes. The manual initialization of feature points limits full automation and may reduce scalability to very large datasets. However, this limitation can be readily mitigated by integrating automatic initialization strategies, e.g., classical keypoint detectors and descriptors (ORB/SIFT/AKAZE) combined with robust matching and RANSAC-based pose estimation, with modern lightweight object/keypoint detectors (YOLO-based or deep keypoint networks) to define regions of interest automatically. Additionally, the current implementation relies on post-processing, and further optimization is required to achieve fully embedded real-time performance, but the first computational performance estimation was also carried out to evaluate the algorithmic overhead with respect to the image acquisition rate. Considering that RGB-D frames are recorded at 15 fps (corresponding to an acquisition period of approximately 0.0667 s per frame), the post-processing algorithm required an average computational time of 0.0188 s per frame when executed on a workstation equipped with an Intel® Core™ i9-13900 (Intel Corporation, Santa Clara, CA, USA). processor and 32 GB of RAM. This means that the processing time accounts for approximately 28% of the image acquisition period, demonstrating that the developed vision pipeline introduces a limited computational load and could achieve real-time feasibility when integrated into embedded hardware. These findings demonstrate the robustness and applicability to field conditions while highlighting the remaining challenges to be addressed in future developments.

6. Conclusions

This work presented the development and experimental test of a vision-based 3D monitoring system for railway freight wagons. The system, based on an RGB-D camera and a computer vision pipeline, was designed to estimate angular excursions and longitudinal displacements of critical coupling components such as hooks and buffers under real operating conditions. The experimental results demonstrated the capability of the proposed approach to track rigid components with good consistency, highlighting correct trends after filtering. Hook angular excursion analysis provided a practical index for identifying potentially dangerous excursions that can lead to hazardous situations, while Buffer displacement monitoring offered an indicator of abnormal elastic deflection of the buffer. These results confirm the feasibility of using vision-based solutions as complementary tools for condition-based maintenance and safety monitoring in freight operations. Despite the promising outcomes, some limitations emerged, particularly those related to illumination variability, image distortions caused by protective housing, and tracking sensitivity to occlusions. The proposed approach demonstrated stable performance and computational feasibility, confirming that the processing pipeline is suitable for real-time operation on embedded systems. The pipeline, thought and designed for embedded execution, showed computational times compatible with real-time operation, which will be experimentally validated in future work. Although the current validation was conducted under real operating conditions, the consistency of the system was indirectly assessed by comparing the measured angular excursions and displacements with reference ranges reported in the literature. This provided a first-level consistency check between the obtained values and known operational limits. Future developments will include laboratory experiments on instrumented prototypes, where the proposed vision-based measurements will be quantitatively compared against a ground-truth reference. These tests will allow the estimation of accuracy and repeatability through standard performance indices such as RMSE and MAE, supporting the definition of quantitative error models for the proposed method. In light of these considerations, the main findings of this research can be outlined as follows: (1) the

proposed RGB-D vision system computes the 3D angular excursion and longitudinal displacement of wagon couplings in real operating conditions; (2) the NCC method, combined with depth reconstruction, achieves accurate and repeatable measurements without the need for physical contact; (3) the experimental results demonstrate consistency with known operational limits reported in the literature, confirming the reliability of the approach; and (4) the developed system architecture ensures real-time data acquisition and computational feasibility for embedded implementation. From a theoretical point of view, this study contributes to extending the application of computer vision and depth-based sensing to 3D geometric monitoring in railway environments, where traditional sensing technologies are difficult to apply. From a practical perspective, the proposed framework supports condition-based maintenance strategies and provides a non-invasive method for the continuous monitoring of freight wagons, potentially enhancing safety and operational efficiency in modern rail transport. The proposed solution is a step toward automated, non-invasive monitoring of mechanical couplings, contributing to the digitalization of freight operations and the improvement of safety, reliability, and efficiency across modern logistics chains. Future work will focus on improving robustness through advanced feature extraction and multi-sensor fusion, coupling the RGB-D stream with inertial and force sensors. Larger datasets and different operational scenarios will be used to further validate the pipeline to assess real-time deployment on embedded hardware.

Author Contributions: Writing (original draft preparation): A.N.; writing (review and editing), conceptualization, methodology, data curation, software, and validation: A.N., S.S., L.L., F.M. and M.T.; review, supervision, and project administration: A.C., L.N., S.D.M., F.M. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the MOST—Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1033 17/06/2022, CN00000023).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Simone Delle Monache was employed by the company Mercitalia Intermodal SpA. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Crozet, Y. Rail freight development in Europe: How to deal with a doubly-imperfect competition? *Transp. Res. Procedia* **2017**, *25*, 425–442. [CrossRef]
2. Djordjević, B.; Ståhlberg, A.; Krmac, E.; Mane, A.S.; Kordnejad, B. Efficient use of European rail freight corridors: Current status and potential enablers. *Transp. Plan. Technol.* **2024**, *47*, 62–88. [CrossRef]
3. Iwnicki, S.; Stichel, S.; Orlova, A.; Hecht, M. Dynamics of railway freight vehicles. *Veh. Syst. Dyn.* **2015**, *53*, 995–1033. [CrossRef]
4. Cheli, F.; Gialleonardo, E.D.; Melzi, S. Freight trains dynamics: Effect of payload and braking power distribution on coupling forces. *Veh. Syst. Dyn.* **2017**, *55*, 464–479. [CrossRef]
5. Ge, X.; Wang, K.; Guo, L.; Yang, M.; Lv, K.; Zhai, W. Investigation on Derailment of Empty Wagons of Long Freight Train during Dynamic Braking. *Shock Vib.* **2018**, *2018*, 2862143. [CrossRef]
6. Liu, P.; Wang, K.; Wang, T. Running Safety of Articulated Freight Wagon Subjected to Coupler Compressive Force. In *ICRT 2017*; ASCE: Reston, VA, USA, 2018; pp. 602–611. [CrossRef]

7. Ulianov, C.; Defossez, F.; Vasić Franklin, G.; Robinson, M. Overview of Freight Train Derailments in the EU: Causes, Impacts, Prevention and Mitigation Measures. In *Traffic Safety*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; Chapter 20, pp. 317–336. [CrossRef]
8. Xie, B.; Chen, S.; Song, P.; Ran, X.; Wang, K. Dynamic identification of coupler yaw angle of heavy haul locomotive: An optimal multi-task ELM-based method. *Mech. Syst. Signal Process.* **2024**, *208*, 110957. [CrossRef]
9. Vale, C.; Mosleh, A.; Mohammadi, M.; Canduco, C.; Ribeiro, D.; Meixedo, A.; Montenegro, P. Smart Rail Infrastructure: Onboard Monitoring with Machine Learning for Track Defect Detection. In Proceedings of the Civil-Comp Conferences-Proceedings of the Sixth International Conference on Railway Technology: Research, Development and Maintenance, Prague, Czech Republic, 1–5 September 2024.
10. Mordia, R.; Verma, A.K. Detection of Rail Defects Caused by Fatigue due to Train Axles Using Machine Learning. *Transp. Infrastruct. Geotechnol.* **2024**, *11*, 3451–3468. [CrossRef]
11. Kumar, S. Advancing Railway Safety through Sensor Fusion and AI-Based Decision Systems. *Int. J. AI Bigdata Comput. Manag. Stud.* **2024**, *5*, 50–59. [CrossRef]
12. Kaiser, I.; Strano, S.; Terzo, M.; Tordela, C. Railway anti-yaw suspension monitoring through a nonlinear constrained estimator. *Int. J. Mech. Control* **2021**, *22*, 35–42.
13. Kaiser, I.; Strano, S.; Terzo, M.; Tordela, C. Anti-yaw damping monitoring of railway secondary suspension through a nonlinear constrained approach integrated with a randomly variable wheel-rail interaction. *Mech. Syst. Signal Process.* **2021**, *146*, 107040. [CrossRef]
14. Bosso, N.; Gugliotta, A.; Magelli, M.; Zampieri, N. Monitoring of railway freight vehicles using onboard systems. *Procedia Struct. Integr.* **2019**, *24*, 692–705. [CrossRef]
15. Bernal, E.; Spiryagin, M.; Cole, C. Onboard Condition Monitoring Sensors, Systems and Techniques for Freight Railway Vehicles: A Review. *IEEE Sens. J.* **2019**, *19*, 4–24. [CrossRef]
16. Lo Schiavo, A. Fully Autonomous Wireless Sensor Network for Freight Wagon Monitoring. *IEEE Sens. J.* **2016**, *16*, 9053–9063. [CrossRef]
17. Moya, I.; Perez, A.; Zabalegui, P.; de Miguel, G.; Losada, M.; Amengual, J.; Adin, I.; Mendizabal, J. Freight Wagon Digitalization for Condition Monitoring and Advanced Operation. *Sensors* **2023**, *23*, 7448. [CrossRef]
18. Kim, H.J.; Byun, Y.S.; Jeong, R.G. Development of Vision-Based Train Positioning System Using Object Detection and Kalman Filter. *J. Electr. Eng. Technol.* **2025**, *20*, 2783–2797. [CrossRef]
19. Schlake, B.W.; Todorovic, S.; Edwards, J.R.; Hart, J.M.; Ahuja, N.; Barkan, C.P. Machine vision condition monitoring of heavy-axle load railcar structural underframe components. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2010**, *224*, 499–511. [CrossRef]
20. Strano, S.; Terzo, M.; Tordela, C. Railway pantograph contact strip monitoring through image processing techniques. In Proceedings of the 2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), Naples, Italy, 13–15 September 2021; pp. 177–181. [CrossRef]
21. Cosenza, C.; Nicolella, A.; Niola, V.; Savino, S. RGB-D Vision Device for Tracking a Moving Target. In Proceedings of the Advances in Italian Mechanism Science, Naples, Italy, 7–9 September 2022; Niola, V., Gasparetto, A., Eds.; Springer: Cham, Switzerland, 2021; pp. 841–848.
22. Cosenza, C.; Nicolella, A.; Esposito, D.; Niola, V.; Savino, S. Mechanical System Control by RGB-D Device. *Machines* **2021**, *9*, 3. [CrossRef]
23. Posada Moreno, A.; Klein, C.; Haßler, M.; Pehar, D.; Solvay, A.; Kohlschein, C. Cargo wagon structural health estimation using computer vision. In Proceedings of the 8th Transport Research Arena, TRA2020, Helsinki, Finland, 27–30 April 2020. [CrossRef]
24. Wang, W.; Ji, T.; Xu, Q.; Su, C.; Zhang, G. A Vision-Based Single-Sensor Approach for Identification and Localization of Unloading Hoppers. *Sensors* **2025**, *25*, 4330. [CrossRef]
25. Mittal, S.; Rao, D. Vision Based Railway Track Monitoring using Deep Learning. *arXiv* **2017**, arXiv:1711.06423. Available online: http://arxiv.org/abs/1711.06423 (accessed on 15 October 2025).
26. Zheng, D.; Li, L.; Zheng, S.; Chai, X.; Zhao, S.; Tong, Q.; Wang, J.; Guo, L. A defect detection method for rail surface and fasteners based on deep convolutional neural network. *Comput. Intell. Neurosci.* **2021**, *2021*, 2565500. [CrossRef]
27. Bai, T.; Gao, J.; Yang, J.; Yao, D. A Study on Railway Surface Defects Detection Based on Machine Vision. *Entropy* **2021**, *23*, 1437. [CrossRef] [PubMed]
28. Pappaterra, M.J.; Pappaterra, M.L.; Flammini, F. A study on the application of convolutional neural networks for the maintenance of railway tracks. *Discov. Artif. Intell.* **2024**, *4*, 30. [CrossRef]
29. Alif, M.A.R.; Hussain, M. Lightweight Convolutional Network with Integrated Attention Mechanism for Missing Bolt Detection in Railways. *Metrology* **2024**, *4*, 254–278. [CrossRef]
30. Zhao, Z.; Kang, J.; Sun, Z.; Ye, T.; Wu, B. A real-time and high-accuracy railway obstacle detection method using lightweight CNN and improved transformer. *Measurement* **2024**, *238*, 115380. [CrossRef]

31. Ferdousi, R.; Laamarti, F.; Yang, C.; Saddik, A.E. A Reusable AI-Enabled Defect Detection System for Railway Using Ensembled CNN. *Appl. Intell.* **2023**, *54*, 9723–9740. [CrossRef]
32. Li, J.; Fu, Y.; Yan, D.; Ma, S.L.; Sham, C.W. An Edge AI System Based on FPGA Platform for Railway Fault Detection. *arXiv* **2024**, arXiv:2408.15245. Available online: http://arxiv.org/abs/2408.15245 (accessed on 15 October 2025).
33. Shaikh, M.Z.; Mehran, S.; Baro, E.N.; Manolova, A.; Uqaili, M.A.; Hussain, T.; Chowdhry, B.S. Design and Development of a Wayside AI-Assisted Vision System for Online Train Wheel Inspection. *Eng. Rep.* **2025**, *7*, e13027. [CrossRef]
34. Gasparini, R.; Pini, S.; Borghi, G.; Scaglione, G.; Calderara, S.; Fedeli, E.; Cucchiara, R. Anomaly detection for vision-based railway inspection. In Proceedings of the European Dependable Computing Conference, Munich, Germany, 7–10 September 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 56–67.
35. Xu, C.; Liao, Y.; Liu, Y.; Tian, R.; Guo, T. Lightweight rail surface defect detection algorithm based on an improved YOLOv8. *Measurement* **2025**, *242*, 115922.
36. Yoo, J.C.; Han, T.H. Fast normalized cross-correlation. *Circuits Syst. Signal Process.* **2009**, *28*, 819–843. [CrossRef]
37. Knap, S.; Suchánek, A.; Harušinec, J. Calculation of the geometric transit through the track curves of a two-section platform wagon. *Transp. Res. Procedia* **2021**, *55*, 853–860. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.