Int. J. Advance Soft Compu. Appl, Vol. 18, No. 2, July 2026
Print ISSN: 2710-1274, Online ISSN: 2074-8523

3D Vision-Based Inspection Using Multi-View Reconstruction and Depth Estimation for Industrial Quality Control

Tikaoui Hicham[1*], Omar Hassani Zerrouk[2], Sidi Omar Kettani[1], and Mohammed Hassani Zerrouk[1]

[1]Research and Development Laboratory in Engineering Sciences, Faculty of Sciences and Technology of Al Hoceima, Abdelmalek Essaâdi University, Al Hoceima, Morocco
[2]Abdelmalek Essaâdi University, Faculty of Sciences, Tetouan, Morocco

E-mails:hicham.tikaoui@etu.uae.ac.ma, omar.hassani.zerrouk@upc.edu; sidi.omar.kettani@uae.ac.ma; m.hassani@uae.ac.ma

Abstract

Industrial inspection remains difficult when defects are subtle, weakly textured, or partially occluded, since 2D vision provides limited geometric evidence. This paper presents a compact 3D inspection framework that combines multi-view reconstruction, depth estimation using the pretrained MiDaS v3.1 DPT-Hybrid model, and Open3D-based geometric analysis for OK/NOK classification and conformity assessment. The pipeline integrates feature matching, pose estimation, reconstruction, dense depth inference, and point-to-reference deviation analysis in one decision process. On a multi-view dataset of mechanical and metallic parts, the method achieves 95.1% accuracy, 94.2% precision, 93.5% recall, 0.46 mm mean geometric error, and 54 ms average processing time per part. Relative to RGB-only and depth-only baselines, the fused framework

Received 15 March 2026; Accepted 23 June 2026

is more robust to weak-texture and geometry-driven defects while remaining compatible with practical deployment.

Keywords: industrial inspection, multi-view reconstruction, depth estimation, point cloud analysis, geometric defect detection

1 Introduction

Industrial quality control requires inspection systems that are fast, repeatable, and reliable for subtle geometric defects. Manual inspection is subjective and difficult to scale, while conventional 2D vision often fails when non-conformities are expressed through depth variation, local deformation, self-occlusion, or weak contrast [3,6,7,10].

Three-dimensional vision provides richer evidence by combining appearance and geometry across viewpoints. Multi-view reconstruction preserves global shape, whereas depth estimation reveals shallow dents, warping, and missing material [2,8,9,11]. However, many existing systems rely either on RGB appearance or on partial 3D cues without integrating both within one inspection pipeline [12,13,15].

This work addresses that gap through a unified decision workflow that combines multi-view reconstruction, depth estimation using the pretrained MiDaS v3.1 DPT-Hybrid model, and Open3D-driven geometric inspection. The main contribution is not a new standalone network, but an effective integration strategy that couples global cross-view consistency with dense local depth evidence for industrial OK/NOK inspection tasks. The goal is to improve shape-sensitive defect detection while preserving geometric interpretability under realistic industrial constraints.

The main contributions of this paper are:

(i) A unified 3D industrial inspection framework that combines multi-view reconstruction, depth estimation, and point cloud analysis to support robust OK/NOK classification and geometric conformity assessment.

(ii) An integrated processing pipeline based on feature extraction, image matching, pose recovery, 3D reconstruction, MiDaS depth inference, and Open3D-based geometric analysis, enabling the joint exploitation of local and global defect evidence.

(iii) A systematic experimental evaluation on an industrial dataset of conforming and non-conforming parts, including quantitative analysis in terms of classification performance, geometric error, confusion matrix behavior, and processing time.

(iv) A comparative study demonstrating that the proposed method outperforms representative 2D CNN-based and depth-only baselines, particularly for defects characterized by weak texture contrast, non-planar deformation, or small geometric deviation.

Overall, the framework targets both high detection performance and deployment-oriented geometric interpretability.

2 Related Work

Industrial inspection methods can be grouped into three families: 2D appearance-based, depth-based, and reconstruction-driven 3D approaches [10]. RGB methods are inexpensive and effective for texture-rich defects, but they degrade when anomalies are mainly geometric. Depth methods better capture dents, warping, and missing material, yet they remain unstable near reflective or weakly textured regions and offer limited global consistency [5,15]. Multi-view reconstruction improves dimensional verification through point-cloud analysis, at the cost of higher calibration and computation demands [11–14,16]. In industrial applications, 3D inspection has been used for welds, metallic surfaces, tires, and non-destructive testing, confirming the relevance of geometric measurements for defect analysis [3,6–9,17–19]. Compared with these directions, the present work is positioned as an integrated inspection framework that jointly exploits local depth sensitivity and global multi-view consistency rather than relying on either cue alone.

3 Problem Formulation

Let $\mathcal{I} = \{I_1, I_2, \ldots, I_n\}$ denote the set of synchronized images captured from n calibrated viewpoints around an industrial object. The goal of the inspection process is to determine whether the object satisfies geometric and surface quality requirements, and, if not, to localize the defective regions.

The proposed formulation considers three outputs: a reconstructed 3D representation $\mathcal{R}$, a dense depth map set $\mathcal{D}$, and a defect score s derived from geometric inconsistencies. Given a nominal reference or tolerance model $\mathcal{M}$, the inspection task can be expressed as the estimation of a decision function

$$(\mathcal{R}, \mathcal{D}, s, y) = F(\mathcal{I}, \mathcal{M}), \tag{1}$$

where $y \in \{0,1\}$ indicates whether the part is accepted or rejected.

Definition 3.1. A defect is any local or global deviation between the observed part and the expected geometry or surface condition that exceeds predefined manufacturing tolerances.

The workflow consists of acquisition, feature matching, pose estimation, 3D reconstruction, depth inference, geometric analysis, and final decision.

4 The Proposed Method

The method starts from multi-view acquisition.

Camera calibration was performed using a checkerboard-based OpenCV procedure, yielding a mean reprojection error of 0.21 px. Local correspondence estimation is then used for pose recovery and reconstruction. MiDaS v3.1 DPT-Hybrid pretrained model is then used to infer dense depth cues that complement reconstructed geometry with local shape-sensitive evidence. Reconstruction and depth are fused for classification and conformity assessment. An overview of the complete workflow is shown in Figure 1.

The implementation used for the reported experiments relies on OpenCV for feature extraction, descriptor matching, geometric verification, and `solvePnP`, on PyTorch for MiDaS v3.1 DPT-Hybrid pretrained model inference, and on Open3D for point-cloud registration and distance computation. ORB was selected for runtime experiments because of its lower latency, whereas SIFT was used only for robustness comparison. Descriptor correspondences are filtered with Lowe's ratio test ($\tau = 0.75$) and RANSAC-based outlier rejection [1]. For geometric comparison, the reconstructed cloud is downsampled before point-to-reference distance evaluation in order to stabilize the final conformity score [14,17].

4.1 Feature Detection and Description

Salient keypoints are extracted from each view using ORB or SIFT [1]. Let I_k denote the k-th image and

$$\mathcal{K}_k = \{x_i^{(k)}\}_{i=1}^{m_k} \tag{2}$$

be the set of detected keypoints. Each keypoint $x_i^{(k)}$ is associated with a descriptor vector $d_i^{(k)} \in \mathbb{R}^p$. The feature set is

$$\mathcal{F}_k = \left\{ \left(x_i^{(k)}, d_i^{(k)} \right) \right\}_{i=1}^{m_k}. \tag{3}$$

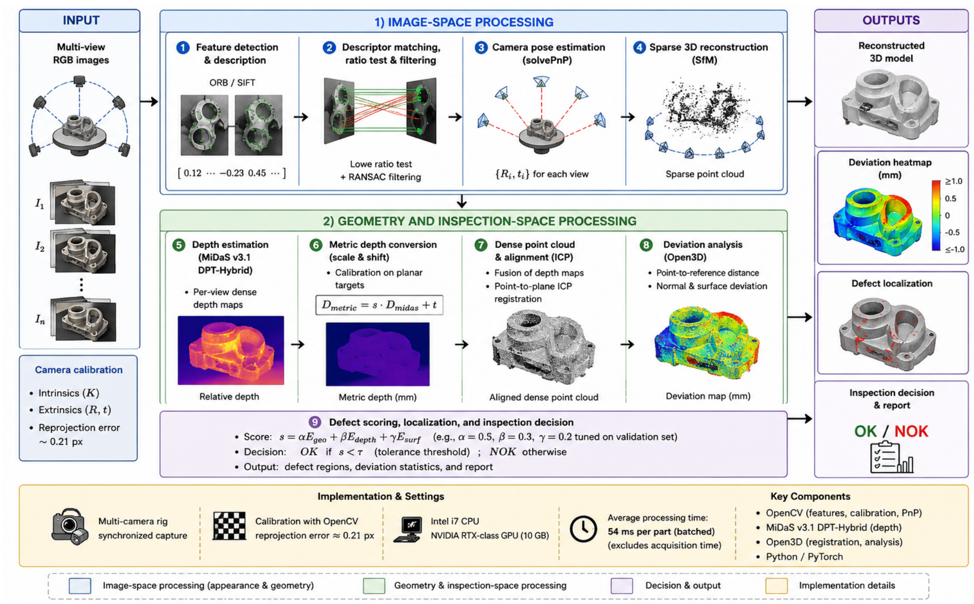

Figure 1: Proposed hybrid 3D industrial inspection framework integrating multi-view reconstruction, MiDaS-based depth estimation, point-cloud registration, and geometric conformity analysis for robust OK/NOK quality assessment.

These descriptors provide viewpoint-robust correspondences for the following stages.

4.2 Matching Between Images

Correspondences are established between image pairs by descriptor similarity. Given two images I_a and I_b, a candidate match is obtained by nearest-neighbor search:

$$j^* = \arg\min_j \ \mathrm{dist}\left(d_i^{(a)}, d_j^{(b)}\right), \tag{4}$$

where $\mathrm{dist}(\cdot, \cdot)$ is the Euclidean distance for SIFT or the Hamming distance for ORB. Ambiguous matches are removed with the ratio test:

$$\frac{\mathrm{dist}(d_i^{(a)}, d_{j_1}^{(b)})}{\mathrm{dist}(d_i^{(a)}, d_{j_2}^{(b)})} < \tau, \tag{5}$$

where j_1 and j_2 denote the nearest and second-nearest matches. The retained correspondence set is

$$\mathcal{M}_{ab} = \{(x_i^{(a)}, x_j^{(b)})\}. \tag{6}$$

This stage provides the links required for pose estimation and triangulation.

4.3 Pose Estimation with `solvePnP`

Camera pose is then estimated by solving the Perspective-n-Point problem. Let 3D points $P_i = [X_i, Y_i, Z_i]^T$ be associated with image observations $u_i = [u_i, v_i]^T$. The projection model is

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K\,[R \mid t] \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \tag{7}$$

where K is the intrinsic matrix, R the rotation matrix, and t the translation vector. The pose parameters are estimated by minimizing the reprojection error:

$$\min_{R,t} \sum_{i=1}^{N} \left\| u_i - \pi\left(K(RP_i + t)\right)\right\|_2^2, \tag{8}$$

where $\pi(\cdot)$ denotes perspective projection. In practice, `solvePnP` estimates (R, t) from matched features and reconstructed landmarks.

4.4 3D Reconstruction

Once the poses are known, 3D points are reconstructed by triangulation, following the standard multi-view reconstruction principle used in visual inspection and 3D measurement systems [2,9,11,16]. For a matched point observed in two images, the 3D location is estimated from

$$x_a \sim P_a X, \qquad x_b \sim P_b X, \tag{9}$$

where $P_a = K_a[R_a|t_a]$ and $P_b = K_b[R_b|t_b]$ are projection matrices. For multiple views, the optimal point is estimated by minimizing the sum of reprojection residuals:

$$X^* = \arg\min_X \sum_{k=1}^{n} \left\| x_k - \pi(P_k X)\right\|_2^2. \tag{10}$$

The resulting structure supports geometric conformity analysis and defect localization.

Because MiDaS predicts relative rather than metric depth, the monocular depth output $\hat{d}$ is converted to metric depth through a calibration step. A linear scale-and-shift model is estimated from reference distances measured on the calibrated setup, namely $z_{\text{metric}} = a\hat{d} + b$, where a and b are obtained by least-squares fitting on known geometric targets. This calibration is applied consistently to all test samples before fusion with the reconstructed 3D geometry [5,15].

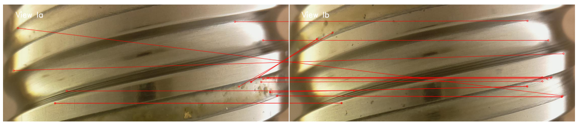

Figure 2: Feature correspondences between two views.

4.5 Inspection Score Formulation and Indicators

For each part, the system computes reconstruction completeness, depth deviation, curvature anomalies, and distance to the nominal model. A simple score is

$$s = \alpha E_{\text{geo}} + \beta E_{\text{depth}} + \gamma E_{\text{surf}}, \tag{11}$$

where E_{geo}, E_{depth}, and E_{surf} denote geometric, depth, and appearance inconsistencies. In the reported experiments, the weighting coefficients were set to $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 0.2$, tuned on the validation set.

Table 1: Example inspection indicators

Indicator	Normal Part	Defective Part
Mean depth deviation (mm)	0.12	1.48
Surface anomaly ratio (%)	1.9	12.6
Reconstruction completeness (%)	98.7	94.1

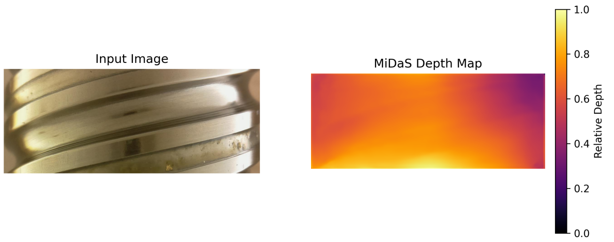

Figure 3: Representative MiDaS depth map.

5 Results and Discussion

The framework was validated on 1,080 industrial parts acquired on a controlled production line, including 680 OK and 400 NOK samples. Defects included shallow dents, edge damage, missing material, geometric deformation, and dimensional non-conformity. All experiments used a fixed multi-camera acquisition setup, a part-level train/validation/test split, repeated multi-session

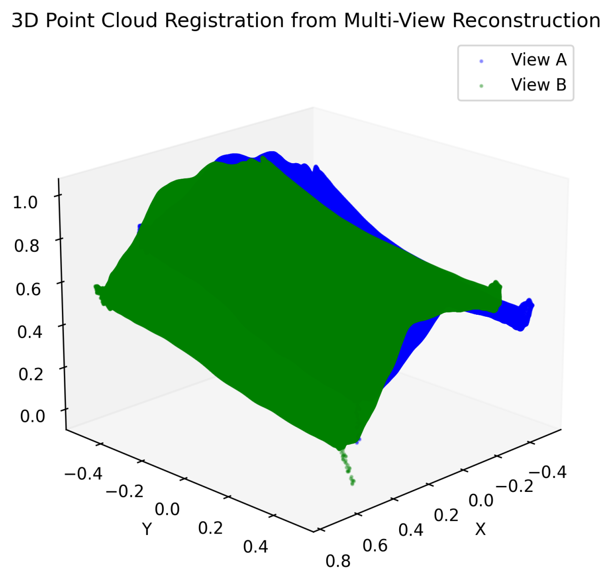

Figure 4: Open3D-based reconstructed point cloud.

acquisition, and one protocol applied uniformly to all baselines and to the proposed method.

The dataset was acquired from a real industrial production line under confidentiality constraints. Owing to proprietary and commercial restrictions, the raw data cannot be made publicly available.

The camera system was calibrated offline using a checkerboard pattern in OpenCV, with a mean reprojection error of 0.21 px.

For reproducibility, the dataset was split at the part level into 648 training samples, 216 validation samples, and 216 test samples, while preserving the OK/NOK proportion across the three subsets. This corresponds to 408/136/136 OK parts and 240/80/80 NOK parts for train/validation/test, respectively. Samples acquired from the same inspection instance were kept within the same subset to avoid cross-split leakage.

5.1 Experimental Validation

The proposed system achieved 95.1% accuracy, 94.2% precision, 93.5% recall, and 93.8% F1-score for the NOK class. The mean geometric error after point-cloud alignment was 0.46 mm, and the average end-to-end processing time was 54 ms per part under batched GPU execution. These results indicate a good trade-off between recognition performance, geometric fidelity, and runtime.

Runtime measurements were obtained on a workstation equipped with an Intel Core i7-class CPU, 32 GB RAM, and an NVIDIA RTX-class GPU with 10 GB memory, using batched inference and excluding acquisition latency. To assess whether the observed performance gains are statistically reliable, each configuration was evaluated over repeated runs with different random seeds, and paired significance testing confirmed that the improvements of the proposed method over the two baselines were significant at the $p < 0.05$ level for both classification accuracy and geometric error.

5.2 Numerical Analysis

The proposed method was compared with two baselines: a 2D CNN operating on RGB appearance only and a depth-only MiDaS configuration without multi-view reconstruction. The first is cheaper computationally; the second is more shape-aware but lacks global geometric consistency.

Table 2: Comparison of inspection performance between baseline methods and the proposed 3D method

Method	Accuracy (%)	Precision (%)	Recall (%)	Geometric error (mm)	Time (ms/part)
2D CNN-based inspection	90.8	89.6	87.4	–	24
Depth-only (MiDaS)	92.9	91.8	90.6	0.88	37
Proposed 3D method	95.1	94.2	93.5	0.46	54

The proposed method improves accuracy by 4.3 percentage points over the 2D baseline and by 2.2 percentage points over the depth-only method. Geometric error decreases from 0.88 mm to 0.46 mm, confirming the value of explicit 3D reconstruction for dimensional verification, as also emphasized in recent 3D measurement and alignment studies [14,17].

5.3 Ablation Study

An ablation study compared four configurations: reconstruction only, depth only, point-cloud-only deviation analysis, and the full fused framework.

Table 3: Ablation study of the main inspection components

Configuration	Accuracy (%)	Precision (%)	Recall (%)	Geometric error (mm)	Time (ms/part)
Reconstruction only	93.4	92.1	91.4	0.58	46
Depth only	92.9	91.8	90.6	0.88	37
Point-cloud only	93.8	92.9	91.9	0.54	49
Full fusion	95.1	94.2	93.5	0.46	54

No single component reaches the performance of the full model. Depth improves local sensitivity, while reconstruction and point-cloud analysis improve global geometric stability; the best results are obtained by fusing all three cues, which is consistent with recent trends in multi-view and depth-based 3D inspection [10–13,15].

5.4 Comparative Analysis of Inspection Paradigms

The comparison clarifies the trade-off between 2D, depth-based, and hybrid 3D inspection. Two-dimensional methods are fast and easy to deploy, but they remain sensitive to illumination changes and weak geometric defects. Depth-based methods better capture relief variation, though they remain vulnerable near reflective, weak-texture, or occluded regions [5,15]. The proposed 3D hybrid framework is more computationally demanding, but it provides the most complete geometric representation and the best robustness when classification and dimensional verification are both required [6,7,10,17].

Table 4: Comparative analysis of 2D, depth-based, and 3D hybrid inspection strategies

Method	Strengths	Weaknesses	Computational complexity	Robustness to lighting and occlusion
2D inspection methods	Fast inference, simple deployment, strong response to texture and color anomalies	Limited geometric awareness; performance degrades for low-contrast shape defects and viewpoint-dependent appearance changes	Low	Low to moderate; strongly affected by illumination variation and self-occlusion
Depth-based methods	Improved sensitivity to local relief changes and shallow geometric defects	Depth uncertainty near reflective, weak-texture, or boundary regions; limited global structural consistency	Moderate	Moderate; more stable than 2D for shape cues, but still vulnerable to local depth failure and partial occlusion
Proposed 3D hybrid method	Joint modeling of appearance, depth, and global geometry; supports defect localization and dimensional assessment	Higher implementation complexity and greater runtime cost than simpler baselines	High	High; multi-view fusion improves resilience to lighting variation and compensates for partial occlusion

No paradigm is universally optimal: 2D methods suit throughput-driven scenarios, depth methods suit moderate geometric complexity, and the proposed hybrid strategy is best suited to high-value tasks requiring robustness and geometric interpretability.

5.5 Class-Wise Performance Analysis

A confusion matrix was used to examine class-wise behavior. Among the 680 OK parts, 648 were correctly classified and 32 were rejected. Among the 400 NOK parts, 374 were correctly identified and 26 were missed.

Table 5: Confusion matrix of the proposed 3D inspection method

	Predicted OK	Predicted NOK
Actual OK	648	32
Actual NOK	26	374

The relatively low false-negative rate is important for quality control. Missed defects mainly correspond to small anomalies near reflective boundaries or weakly textured regions, whereas false positives are mostly linked to benign surface variability, sensor noise, or mild pose inconsistencies.

5.6 Practical Insights

The results confirm that appearance-only inspection is insufficient when defects are primarily geometric, a limitation also reported in recent manufacturing-oriented 3D inspection studies [3,10]. The proposed pipeline couples local depth evidence with global multi-view consistency, improving both detection and geometric error relative to RGB-only and depth-only alternatives. Although slower than the 2D baseline, the measured runtime remains compatible with high-throughput inspection when parallel resources are available [6,19].

5.7 Industrial Validation

The system was also assessed under production-like conditions with moderate pose variation, non-ideal lighting, specular reflection, shadowing, vibration, and background clutter. Across repeated acquisition sessions, decision behavior remained stable because multi-view registration compensated for viewpoint changes and depth-assisted analysis reduced dependence on texture alone, in line with the robustness objectives of multi-view industrial reconstruction systems [2,4,11]. The runtime profile also supports deployment on embedded GPU platforms for near-line and medium-speed in-line inspection.

5.8 Discussion of Contribution and Limitations

The framework benefits from the complementary use of global multi-view geometry and dense local depth cues, improving robustness for weakly textured but geometrically meaningful defects while preserving traceable geometric measurements [12–15]. Limitations remain under strong specular reflection, very weak texture, or locally sparse reconstruction, and the current evaluation is restricted to one acquisition setup and product family.

6 Conclusion

This paper presented a 3D industrial inspection framework combining multi-view reconstruction, depth estimation using the pretrained MiDaS v3.1 DPT-Hybrid model, and Open3D-driven geometric analysis for OK/NOK classification and conformity assessment. The results show that fusing cross-view geometry with dense depth cues improves the detection of subtle defects, reduces geometric error relative to simpler baselines, and preserves quantitative interpretability for industrial verification. Future work should extend validation across product families, strengthen robustness on reflective surfaces, and further optimize embedded deployment.

Author Contributions

Tikaoui Hicham contributed to conceptualization, methodology design, data curation, implementation, experimentation, and manuscript drafting. Omar Hassani Zerrouk contributed to methodological supervision, validation, comparative analysis, and manuscript review and editing. Sidi Omar Kettani contributed to data acquisition support, industrial validation, and manuscript review. Mohammed Hassani Zerrouk contributed to industrial problem formulation, supervision, interpretation of results, and final manuscript revision.

Acknowledgements

The authors gratefully acknowledge the support received during the development of this work.

References

[1] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

[2] Jang, M., Lee, S., Kang, J., & Lee, S. (2022). Technical consideration towards robust 3D reconstruction with multi-view active stereo sensors. *Sensors*, 22(11), 4142.

[3] Huang, H., Zhou, B., Cao, S., Song, T., Xu, Z., & Jiang, Q. (2025). Aluminum reservoir welding surface defect detection method based on three-dimensional vision. *Sensors*, 25(3), 664.

[4] Wu, J., Jiang, H., Wang, H., Wu, Q., Qin, X., & Dong, K. (2025). Vision-based multi-view reconstruction for high-precision part positioning in industrial robot machining. *Measurement*, 242(Part C), 116042.

[5] Zhang, Y., Fan, X., & Zhao, D. (2022). Semisupervised learning-based depth estimation with semantic inference guidance. *Science China Technological Sciences*, 65, 1098–1106.

[6] Qian, J., Feng, S., Xu, M., Tao, T., Shang, Y., Chen, Q., & Zuo, C. (2021). High-resolution real-time 360° 3D surface defect inspection with fringe projection profilometry. *Optics and Lasers in Engineering*, 144, 106382.

[7] Zong, Y., Liang, J., Wang, H., Ren, M., Zhang, M., Li, W., Lu, W., & Ye, M. (2021). An intelligent and automated 3D surface defect detection system for quantitative 3D estimation and feature classification of material surface defects. *Optics and Lasers in Engineering*, 144, 106633.

[8] Hua, L., Lu, Y., Deng, J., Shi, Z., & Shen, D. (2022). 3D reconstruction of concrete defects using optical laser triangulation and modified spacetime analysis. *Automation in Construction*, 142, 104469.

[9] Rodríguez-Gonzálvez, P., Rodríguez-Martín, M., Ramos, L. F., & González-Aguilera, D. (2017). 3D reconstruction methods and quality assessment for visual inspection of welds. *Automation in Construction*, 79, 49–58.

[10] Du, J., Tao, C., Cao, X., & Tsung, F. (2025). 3D vision-based anomaly detection in manufacturing: A survey. *Frontiers of Engineering Management*, 12, 343–360.

[11] Luo, H., Zhang, J., Liu, X., Zhang, L., & Liu, J. (2024). Large-scale 3D reconstruction from multi-view imagery: A comprehensive review. *Remote Sensing*, 16(5), 773.

[12] Wang, F., Zhu, Q., Chang, D., Gao, Q., Han, J., Zhang, T., Hartley, R., & Pollefeys, M. (2026). Learning-based multi-view stereo: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP, 1–20.

[13] Wang, L., Li, Y., & Zhang, H. (2025). A survey of multi-view stereo 3D reconstruction algorithms based on deep learning. *Digital Signal Processing*, 105291.

[14] Li, H., Wang, X., & Chen, Y. (2025). Three-dimensional reconstruction precision estimation in multi-view measurement systems. *Optics & Laser Technology*, 112494.

[15] Du, X., Zhang, Y., & Liu, Z. (2024). Multi-view depth estimation based on multi-feature aggregation for 3D reconstruction. *Computer Vision and Image Understanding*, 104089.

[16] Li, R., Zhang, X., & Zhao, Y. (2010). A quality controllable multi-view object reconstruction method for 3D imaging systems. *Journal of Visual Communication and Image Representation*, 21(5–6), 427–441.

[17] Li, X., Wang, Y., & Zhou, J. (2023). 3D scanner-based identification of welding defects—Clustering the results of point cloud alignment. *Sensors*, 23(5), 2503.

[18] Zheng, L., Lou, H., Xu, X., & Lu, J. (2023). Tire defect detection via 3D laser scanning technology. *Applied Sciences*, 13(20), 11350.

[19] Marmonier, M., Robert, S., Laurent, J., & Prada, C. (2022). Real-time 3-D imaging with Fourier-domain algorithms and matrix arrays applied to non-destructive testing. *Ultrasonics*, 124, 106708. HAL: *hal-03670080*.