Human Systems Engineering and Design (IHSED2025), Vol. 198, 2025, 289–298
https://doi.org/10.54941/ahfe1006808

Automated Procedural Error Detection in Human-Robot Collaborative Assembly Using Vision-Based Template Matching

Isaiah Nassiuma, Ella-Mae Hubbard, and Yee Mey Goh

Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, UK

ABSTRACT

In collaborative human-robot assembly, robust error identification is paramount for ensuring process integrity and safety, particularly in the post task phase where a comprehensive analysis provides an opportunity to identify subtle and cumulative errors that may have been missed in real-time. Traditional manual verification is often tedious and prone to human error, including oversight and fatigue, which can compromise quality. This paper evaluates the efficacy of an automated, vision-based error detection system using OpenCV template matching as a more reliable alternative. Our method identifies procedural errors, such as missed components or out-of-sequence operations, by comparing real-time images of the assembly state against a library of reference templates that depict correctly completed procedural steps. Visual dissimilarity metrics are used to automatically flag deviations from the expected sequence. Experimental results demonstrate that the automated system significantly outperforms manual verification in the consistent and rapid identification of both missing and mis-sequenced assembly steps. Whilst its performance can be influenced by challenges such as variable lighting and low-contrast features, the vision-based approach proved substantially more dependable than human inspection especially for structured and defined tasks where the objects consistent and predicted visual features. We conclude that template matching provides a robust and scalable solution for quality control in collaborative assembly tasks. This automated approach enhances operational efficiency and safety, though further tuning may be required to optimise performance in visually complex environments.

Keywords: Human-robot collaboration, Template matching, Error detection, Computer vision

INTRODUCTION

Human-robot collaboration has increased in manufacturing due to their ability to combine human flexibility with robotic precision (Fan, Zheng and Li, 2022). These systems are particularly valuable in complex assembly tasks where human dexterity is essential, but consistency and repeatability are challenging to maintain (Gervasi *et al.*, 2023). However, despite advancements in robotic automation, human involvement introduces variability, leading to procedural errors such as missing components, incorrect sequencing, or misalignments (Caterino *et al.*, 2023).

© 2025. Published by AHFE Open Access. All rights reserved.

As robots become increasingly integrated into manufacturing, much attention has been given to ensuring their safe and accurate performance during collaboration. Robotic systems typically follow pre-programmed routines or AI-driven algorithms, but human involvement may introduce unpredictable deviations from expected workflows (Lodhi and Zeb, 2025). To maintain process integrity without compromising worker productivity, there is a growing need for automated, adaptive error detection systems (Caterino *et al.*, 2023).

To ensure process integrity in manufacturing, standard procedures and instructions are typically enforced, supported by quality control personnel who conduct manual inspections during or after assembly. However, these manual checks are susceptible to human limitations such as fatigue, oversight, and cognitive bias (Kim *et al.*, 2020). Errors detected post-assembly can result in costly rework or scrapping of components, highlighting the need for earlier intervention.

To address these challenges, vision-based quality control systems have emerged as effective, non-intrusive solutions for real-time monitoring. These systems offer the advantage of continuous oversight without interrupting the workflow. While advanced methods such as deep learning and stereo vision have demonstrated strong performance in detecting defects, their practical deployment is often hindered by high computational requirements and the need for extensive training datasets (Frustaci *et al.*, 2022). These limitations restrict their scalability, particularly in resource-constrained environments.

In dynamic human-robot collaborative environments, errors are not static defects but transient events that occur during the workflow (Puttero *et al.*, 2023). Detecting these in process is essential. A promising approach involves using low-cost, computationally efficient methods such as template matching (Duan *et al.*, 2024). This technique is particularly suited for structured environments where assembly steps are well-defined, offering the potential for immediate feedback on deviations and significantly enhancing error prevention during the process itself.

RELATED WORK

In this section we shall look at the problem from three areas since all these areas combined provide the unique approach this paper has taken. We will start with human–robot collaboration, error identification, and lastly computer vision object detection.

Previous work on human-robot collaboration has highlighted the importance of reviewing the mistakes in assembly tasks (Antonelli and Stadnicka, 2019), whether they're from a human (Caterino *et al.*, 2023) or by a robot (Stiber, Taylor and Huang, 2022). Two areas have had a substantial amount of research done on them, one being the understanding of how humans react to and /or resolve robot errors (Honig and Oron-Gilad, 2018; Liu and Wang, 2021; Stiber, Taylor and Huang, 2022). Secondly, we have the human errors based on their actions that might be unsafe as they interact with a robot. The insight provided by the scope of human errors and the

type of errors has been explored (Liu *et al.*, 2025). However, most of these errors are based on what the human does in regard to their safety around the robot and not particularly their role in the assembly process. Research shows human errors affect the efficiency, safety, and performance of a system (Esposito *et al.*, 2025). One of the highlighted errors, amongst other errors, is a sequence error, which entails deviating from a predefined sequence through insertion, omission, substitution, or reversal of an action (Klages, Graf and Zaeh, 2024; Esposito *et al.*, 2025).

Various computer vision techniques have been used to analyse the errors, including object detection, which involves training a dataset (Conati *et al.*, 2020). Template matching used in this scenario is based on 2D images with a fixed camera, as it reduces computational cost as compared to 3D images. Action and task recognition have been used to check for procedural errors (Conati *et al.*, 2020), but the algorithm needs to be trained before deployment. For example, Bovo *et al.* (2020) used action recognition by detecting the errors through recognising hand movements and eye gaze, segmenting video frames where a person places an object. Another approach is one taken by Zhang *et al.* (2022) used recurrent neural network to detect the faults through the video recordings. Both action recognition and object recognition commonly utilise RNNs (Soran, Farhadi and Shapiro, no date; Ay and Emel, 2025).

In terms of assembly verification, template matching has proven effective for tracking part assembly at each stage, as demonstrated by Pang *et al.* (2023). It has also been useful for identifying object-related errors (Kong, Wu and Song, 2022). When the assembly process is standardised, meaning object orientation, lighting, and positioning are consistent, template matching becomes a convenient and reliable method for part identification. Pang *et al.* (2023) provide the foundational viability of the method but also indicates that the error classification and the integration of action-task verification could further enhance its effectiveness.

METHODOLOGY

Experimental Setup

The experiment aimed to evaluate the effectiveness of an automated, vision-based error detection system during a collaborative human-robot LEGO® assembly task. In this setup (see Figure 1), a robot constructed the base of three sequential LEGO® structures, while a human participant completed the top layers using predefined instructions. The assemblies had to be completed in a strict order—assembly 1 before 2, and 2 before 3. Each structure had a unique configuration, clearly depicted in printed reference images that were available to the participant throughout the task. The entire process was continuously recorded by an overhead camera for later analysis. Each participant completed the assemblies individually, and their progression from one assembly to the next was used as an indicator that the previous structure had been completed.

Figure 1: Experiment setup.

Error Categories

Three distinct categories of errors were defined for detection as shown in Table 1.

Table 1: Error categories.

Error	Description
Sequence Violation	Commencing a subsequent assembly before the prior one was fully and correctly completed.
Construction Error	Assembling components incorrectly, including the misplacement, omission, or incorrect orientation of LEGO® pieces relative to the reference design.
Mixed Error	Any instance where both a sequence violation and one or more construction errors occurred simultaneously.

Automated Error Detection System

The video recordings were processed by a custom error detection pipeline developed using the OpenCV template matching library (OpenCV: Template Matching, 2025). The system employed template matching to identify deviations between the assembly produced by each participant and a library of reference images corresponding to correctly completed configurations. Video frames were periodically extracted and compared against these templates to detect discrepancies.

Dissimilarity metrics were calculated using normalised cross-correlation referred to as cv2.TM_CCORR_NORMED (OpenCV: Object Detection, 2025). Deviations exceeding pre-defined thresholds triggered an automatic error flag. The system was designed for minimal human intervention and

processed frames at an average rate of under two seconds each, enabling near-real-time analysis.

Establishing a Ground Truth via Human Verification

To create a robust benchmark for evaluating the automated system, all video recordings were meticulously analysed by a single, independent human reviewer. This reviewer was an expert in the assembly task but had no involvement in the experiment's execution.

To ensure the credibility and consistency of the human-generated labels, a formal intra-rater reliability protocol was implemented (see Figure 2). First, a detailed rubric was created, providing objective, unambiguous criteria for classifying each type of error. Secondly, the reviewer analysed all video recordings and logged any observed errors according to the rubric. This was followed by a four-week "washout" period, after which the reviewer had no direct recollection of their initial specific judgements. Finaly, the reviewer was re-analysing the videos using the same rubric. The consistency between the first and second assessment passes was then quantified to validate the reliability of the reviewer's judgements.

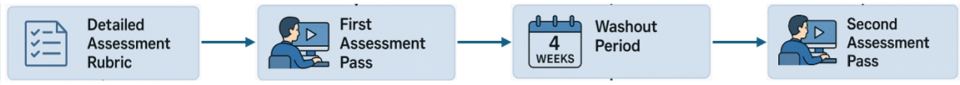

Figure 2: Intra-rater reliability protocol for manual verification.

Comparison Framework and Metrics

The reliability of the expert reviewer's assessments was confirmed by calculating Cohen's Kappa (κ) on the two passes, which indicated a high level of intra-rater agreement (κ>0.65). Having established the reviewer's high consistency, the assessments from their first pass were adopted as the validated ground truth for the experiment.

The verification speed which is the average time taken by the automated system to analyse a single assembly, compared with the time taken by the human reviewer. Additionally, the performance of the automated system was then evaluated against this ground truth using the metrics described in Figure 3 and Table 2.

Table 2: Metrics used to evaluate the performance of the automated system.

Metric	Definition	Formula
Specificity	Proportion of actual negatives correctly identified.	TN / (TN + FP)
Sensitivity/ Recall	Proportion of actual positives correctly identified.	TP / (TP + FN)
Precision	Proportion of predicted positives that are actually positive.	TP / (TP + FP)
F1 Score	Harmonic mean of precision and recall.	2 *(Precision * Recall) / (Precision + Recall)

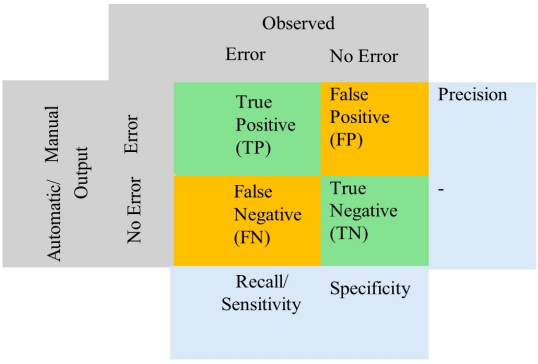

Figure 3: Confusion matrix of true and false negatives and positives.

RESULTS AND DISCUSSION

The performance of the automated verification system was evaluated against manual verification across the three distinct error types using key metrics such as recall, precision, specificity, and F1 score to determine their effectiveness in identifying and classifying these errors (See Figure 4).

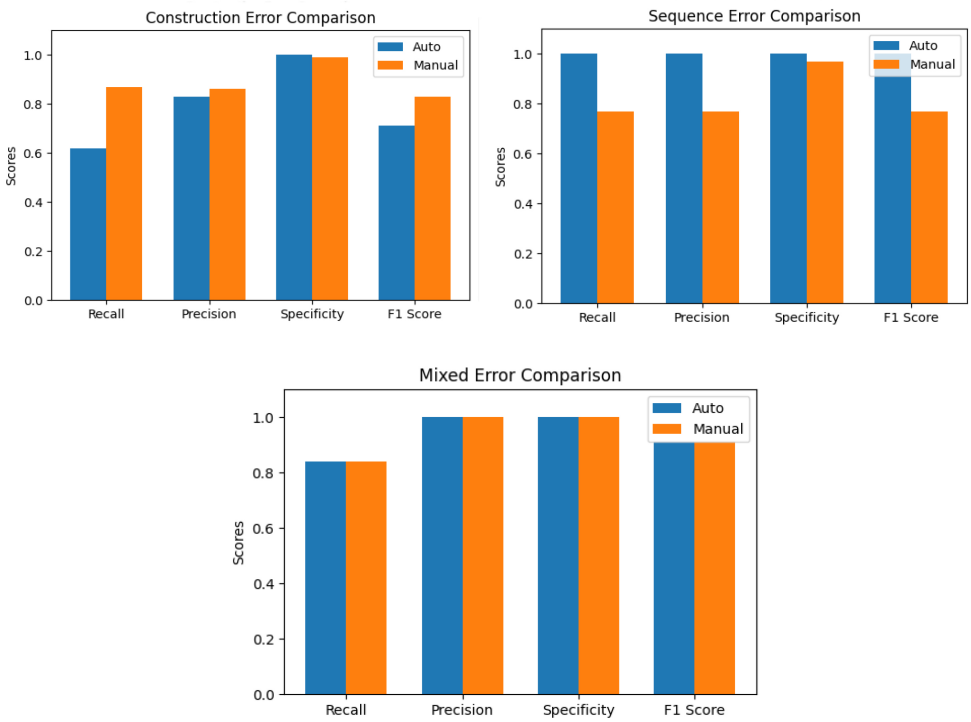

Figure 4: The three errors and the graphs of the metrics based on their performance.

In the construction category, the manual method outperformed the automated method achieving a higher recall of 0.875 and an F1 score of 0.875 compared to that of the automated system, 0.625 and 0.714, respectively.

Both methods demonstrated equal specificity of 0.990, indicating a comparable ability to correctly identify error-free assemblies. Observational analysis of the video recordings revealed that the automated method misclassified items with subtle differences that could not be identified by the template matching algorithm and where the lighting and background of the template had changed. This shows that while the algorithm is effective in general pattern recognition, it may lack sensitivity to fine-grained structural deviations. Adjusting the algorithm's accuracy threshold could improve detection but must be carefully balanced to avoid misclassifying valid assemblies affected by orientation, lighting, or scale variations.

For sequence errors, the automated method demonstrated perfect performance, with recall, precision, specificity, and F1 score all at 1.000. This indicates it was able to detect every sequence error without any false positives or negatives. The manual method, however, showed weaker performance, with recall and precision both at 0.765, and a matching F1 score of 0.765. The manual analysis while rapidly going through the video would have led to this. This also might be introduced by fatigue since once a person is looking at several videos, they will tend to get tired of doing that every time. Here the reliance on the automated checks might be more useful than the manual verification. Finally, in the missed errors category both methods performed similarly showing no difference in identifying areas where both construction and sequence errors occurred. However, they both missed identification at different points hence the recall of 0.833.

The overall performance comparison between the two methods across all observations reveals a consistent advantage in favour of the automated approach across all evaluated metrics (See Figure 5).

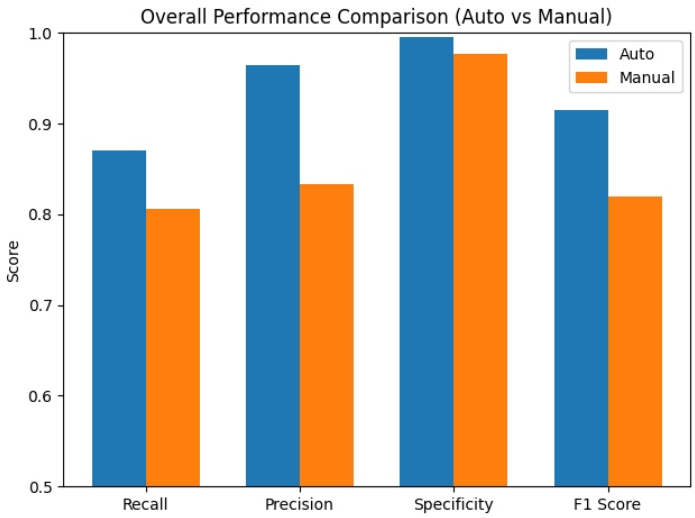

Figure 5: Overall comparison of the automated and manual methods.

The automated method achieved higher recall (0.871 vs. 0.806), precision (0.964 vs. 0.833), specificity (0.995 vs. 0.977), and F1 score (0.915 vs. 0.820), indicating greater consistency and reliability in detecting and

classifying assembly errors. These results show that the template matching method may be better suited for high-throughput environments requiring accurate and consistent error detection. However, statistical analysis using the Wilcoxon Signed-Rank Test revealed that the difference in F1 scores was not statistically significant ($p = 0.655$), implying that the observed performance advantage may not be conclusive given the limited sample size.

Verification speed was evaluated for both manual and automated methods under controlled playback conditions. Initially, the average duration required to manually review the video recordings at double-speed playback was calculated to be 128 seconds. Under the same playback conditions, the automated system completed verification in 98 seconds, demonstrating a faster processing capability. This suggests that the optimized automated method can outperform human reviewers when operating under equivalent time constraints. A second comparison assessed actual verification times under typical usage conditions used during the experiment. Manual reviewer, who was permitted to skip through footage, achieved an average verification time of 34 seconds per video. In contrast, the automated system, capable of parallel processing 8 recordings at a time, reduced the average verification time to 26 seconds per video. The results demonstrate the superior overall performance of the template matching method to the manual verification under this experimental setup.

CONCLUSION

The evaluation demonstrates that while the automated system shows promise for assembly verification tasks, background consistency is essential for reliable real-world deployment. The template matching proved to be as good overall, including both precision and recall in. This, therefore, would be ideal where workstations have a fixed background, and the work surfaces are always similar to those taken during the template capture. In real-time scenarios that have the same background, the automated system would be more practical due to its speed in identifying the objects; however, if the condition or background changes, it would be better to have a human verifier analyse the results and mostly the false negatives in sequence and construction. This work lays the groundwork for future exploration of a hybrid approach, where the automated system operates with a high accuracy threshold, particularly for construction errors, and flags only potentially faulty items for human review. This targeted verification could even be further streamlined by highlighting specific frames where anomalies are detected, allowing reviewers to quickly assess critical moments without scanning entire videos. Such a system could offer a scalable, efficient, and reliable solution for real-world assembly verification tasks.

REFERENCES

Antonelli, D. and Stadnicka, D. (2019) 'Predicting and preventing mistakes in human-robot collaborative assembly', *IFAC-PapersOnLine*, 52(13), pp. 743–748. Available at: https://doi.org/10.1016/J. IFACOL.2019.11.204.

Ay, O. and Emel, E. (2025) 'Real-Time Assembly Task Validation Using Deep Learning-Based Object Detection and Operator's Hand-Joints Trajectory Classification', *IEEE Access*, 13, pp. 57009–57029. Available at: https://doi.org/10.1109/ACCESS.2025.3554263.

Bovo, R. *et al.* (2020) 'Detecting Errors in Pick and Place Procedures Detecting Errors in Multi-Stage and Sequence-Constrained Manual Retrieve-Assembly Procedures'. Available at: https://doi.org/10.1145/3377325.3377497.

Caterino, M. *et al.* (2023) 'A Human Error Analysis in Human–Robot Interaction Contexts: Evidence from an Empirical Study', *Machines 2023, Vol. 11,* 11(7), p. 670. Available at: https://doi.org/10.3390/MACHINES11070670.

Conati, C. *et al.* (2020) 'Comparing and Combining Interaction Data and Eye-tracking Data for the Real-time Prediction of User Cognitive Abilities in Visualization Tasks', *ACM Transactions on Interactive Intelligent Systems*, 10(2), p. 12. Available at: https://doi.org/10.1145/3301400.

Duan, J. *et al.* (2024) 'HRC for dual-robot intelligent assembly system based on multimodal perception', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 238(4), pp. 562–576. Available at: https://doi.org/10.1177/09544054231167209/ASSET/CD7C8D3D-FE41-4611-B4A6-AD44C37E25CC/ASSETS/IMAGES/LARGE/10.1177_09544054231167209-FIG15. JPG.

Esposito, C. *et al.* (2025) 'The role of human error in human robot interaction', *Procedia Computer Science*, 253, pp. 2347–2357. Available at: https://doi.org/10.1016/J. PROCS.2025.01.295.

Fan, J., Zheng, P. and Li, S. (2022) 'Vision-based holistic scene understanding towards proactive human–robot collaboration', *Robotics and Computer-Integrated Manufacturing*, 75, p. 102304. Available at: https://doi.org/10.1016/J. RCIM.2021.102304.

Frustaci, F. *et al.* (2022) 'Robust and High-Performance Machine Vision System for Automatic Quality Inspection in Assembly Processes', *Sensors 2022, Vol. 22,* 22(8), p. 2839. Available at: https://doi.org/10.3390/S22082839.

Gervasi, R. *et al.* (2023) 'Manual assembly and Human–Robot Collaboration in repetitive assembly processes: a structured comparison based on human-centered performances', *International Journal of Advanced Manufacturing Technology*, 126(3–4), pp. 1213–1231. Available at: https://doi.org/10.1007/S00170-023-11197-4/TABLES/7.

Honig, S. and Oron-Gilad, T. (2018) 'Understanding and resolving failures in human-robot interaction: Literature review and model development', *Frontiers in Psychology*, 9(JUN), p. 351644. Available at: https://doi.org/10.3389/FPSYG.2018.00861/BIBTEX.

Kim, S. K. *et al.* (2020) 'Errors in Human-Robot Interactions and Their Effects on Robot Learning', *Frontiers in Robotics and AI*, 7, p. 558531. Available at: https://doi.org/10.3389/FROBT.2020.558531/BIBTEX.

Klages, B., Graf, J. and Zaeh, M. (2024) 'Human errors in manual assembly – A survey on current and future relevance', *Procedia CIRP*, 130, pp. 1556–1561. Available at: https://doi.org/10.1016/J. PROCIR.2024.10.282.

Kong, Q., Wu, Z. and Song, Y. (2022) 'Online detection of external thread surface defects based on an improved template matching algorithm', *Measurement*, 195, p. 111087. Available at: https://doi.org/10.1016/J. MEASUREMENT.2022.111087.

Liu, H. and Wang, L. (2021) 'Collision-free human-robot collaboration based on context awareness', *Robotics and Computer-Integrated Manufacturing*, 67, p. 101997. Available at: https://doi.org/10.1016/J. RCIM.2020.101997.

Liu, L. *et al.* (2025) 'Human Error Identification and Risk Prioritization in Human–Robot Collaboration in Manufacturing', *Human Factors and Ergonomics in Manufacturing & Service Industries*, 35(3), p. e70012. Available at: https://doi.org/10.1002/HFM.70012.

Lodhi, S. K. and Zeb, S. (2025) 'AI-Driven Robotics and Automation: The Evolution of Human-Machine Collaboration', *Journal of World Science*, 4(4), pp. 422–437. Available at: https://doi.org/10.58344/JWS. V4I4.1389.

OpenCV: Object Detection (2025). Available at: https://docs.opencv.org/3.4/df/dfb/group__imgproc__object.html#gga3a7850640f1fe1f58fe91a2d7583695dab65c042ed62c9e9e095a1e7e41fe2773 (Accessed: 8 August 2025).

OpenCV: Template Matching (2025). Available at: https://docs.opencv.org/3.4/d4/dc6/tutorial_py_template_matching.html (Accessed: 8 August 2025).

Pang, J. *et al.* (2023) 'A verification-oriented and part-focused assembly monitoring system based on multi-layered digital twin', *Journal of Manufacturing Systems*, 68, pp. 477–492. Available at: https://doi.org/10.1016/J. JMSY.2023.05.008.

Puttero, S. *et al.* (2023) 'Towards the modelling of defect generation in human-robot collaborative assembly', *Procedia CIRP*, 118, pp. 247–252. Available at: https://doi.org/10.1016/J. PROCIR.2023.06.043.

Soran, B., Farhadi, A. and Shapiro, L. (no date) 'Generating Notifications for Missing Actions: Don't forget to turn the lights off!'

Stiber, M., Taylor, R. and Huang, C. M. (2022) 'Modeling Human Response to Robot Errors for Timely Error Detection', *IEEE International Conference on Intelligent Robots and Systems*, 2022-October, pp. 676–683. Available at: https://doi.org/10.1109/IROS47612.2022.9981726.

Zhang, Z. *et al.* (2022) 'Real-Time Human Fault Detection in Assembly Tasks, Based on Human Action Prediction Using a Spatio-Temporal Learning Model', *Sustainability 2022, Vol. 14,* 14(15), p. 9027. Available at: https://doi.org/10.3390/SU14159027.