Skill-Based Adaptation through Intuitive Interfaces: Multi-Modal Guidance Systems for Industrial Environments

Sharath Chandra Akkaladevi[1[0000-0001-9105-6269]], Johann J. Mitteramskogler[1[0000-0003-3332-5461]], Gerhard Ebenhofer[1[0009-0008-3521-5972]], and Andreas Pichler[1[0000-0003-4312-2369]]

[1] PROFACTOR GmbH, Im Stadtgut D1, 4407 Steyr-Gleink, Austria

Abstract. High-mix, low-volume manufacturing environments, particularly in the automotive sector, require precise defect detection and rework processes to maintain quality standards while accommodating workforce variability. Traditional training methods often rely on physical tools and static instructions, which hinder efficiency and limit adaptability across different skill levels. Addressing these challenges, this research introduces a multi-modal guidance system that supports both novice and expert workers by providing real-time, skill-adaptive assistance. The proposed system integrates camera-based projectors and interactive graphical user interfaces (GUIs) to deliver intuitive, dynamic guidance tailored to user expertise. Operating in two distinct modes, Novice Mode enables trainees to mark defects on sub-assembled parts using real-time hand tracking, with projected visual feedback eliminating the need for physical markings. Expert Mode projects defect locations and rework instructions directly onto parts, complemented by GUI-based insights for precise corrections. Laboratory tests across various parts and approximately 10 defect types demonstrate the system's effectiveness in improving training outcomes and streamlining quality control processes. Key results highlight enhanced efficiency, repeated use of training parts, and improved user engagement through features such as dynamic interaction, 3D model integration, and gamification. The modular and scalable design of the system lays a foundation for intelligent workplace assistance, with future implications for broader adoption in diverse industrial settings to boost productivity and adaptability.

Keywords: Multi-modal guidance, workplace assistance, skill-based adaptation, intuitive interfaces, industrial training, hand tracking, projection systems.

1 Introduction

High-mix, low-volume (HMLV) manufacturing, particularly in the automotive sector, demands high adaptability and precision in identifying and reworking defects such as scratches or dents on semi-assembled components. This labor-intensive process relies on skilled operators and effective training systems to maintain quality and consisten-

cy. As workforce skill levels vary, adaptive guidance systems are essential to support both novice and experienced workers in inspection and rework tasks.
Traditional training methods, using physical marking tools and printed instructions, increase hardware overhead, limit part reuse due to permanent markings, and provide static guidance regardless of user proficiency. These limitations hinder efficiency and slow down skill development.

To address these challenges, this research introduces a multi-modal guidance system that adapts dynamically to user expertise. The system operates in two modes: **Novice Mode** enables workers to identify and mark defects using real-time hand tracking, with projected feedback eliminating physical markings; **Expert Mode** assists experienced workers by projecting defect locations and rework instructions directly onto parts, supported by contextual insights via an interactive GUI.

The proposed system offers several key contributions:

- Adapting hand-tracking methods [1] for real-time defect detection, facilitating natural and marker-free user interaction.
- Projection-based feedback for novice and expert workers, with tailored instructional content and visual cues.
- Integration of skill-based adaptation and gamification elements, designed to enhance engagement, accelerate skill development, and standardize rework procedures.

By unifying training and rework assistance into a flexible, scalable platform, the system offers an efficient solution for quality assurance and rapid worker onboarding in industrial environments.

2 State of the Art

Modern HMLV manufacturing increasingly leverages adaptive, multi-modal human-machine interfaces (HMIs) to support operators in complex inspection and rework tasks. These interfaces combine projection-based guidance, augmented reality (AR), graphical user interfaces (GUIs), and natural interaction methods such as gesture or voice input.

Adaptive interfaces adjust content based on user skill and task context, reducing cognitive load and errors. Projection systems and AR overlays have been shown to improve inspection accuracy and user engagement (Seeliger et al. [2]; Havlíková et al. [3]). Suzuki et al. [4] introduced a taxonomy for designing adaptive AR systems in human-robot interaction. **Spatial augmented reality** (SAR) further supports task execution and ergonomic safety through visual cues (Wedral et al. [5]), while pose estimation enhances personalized, real-time feedback (Stübl et al. [9], Ikeda et al. [10]). These adaptive tools help deliver context-aware assistance as users interact with workpieces. **Skill-based systems** are replacing static, paper-based training with interactive, modular environments. AR-based frameworks provide step-by-step guidance and real-time correction, significantly reducing onboarding time (Stübl et al. [8]). In spot repair scenarios, VR tools with intuitive modalities, such as digital pens, improve performance and satisfaction (Puthenkalam et al. [6]). **Multi-modal systems** integrate various interfaces—projections, AR, GUIs, and gestures—to support diverse tasks. XR platforms scale SAR across factory settings (Pönitz et al. [7]), and gesture-based control enhances interaction without wearable hardware (Stübl et al. [9]). These systems are effective in tasks from defect annotation to rework visualization.

2.1 Comparison with Existing Systems

Traditional training and defect detection systems are manual, static, and inflexible, often leading to inconsistency and high training costs. In contrast, skill-adaptive, real-time systems provide more consistent and efficient support. Projection-based interfaces reduce errors and improve throughput (Stübl et al. [8]), while gesture-controlled projections enhance engagement and safety compliance (Wedral et al. [5]).

These findings underline the shift toward modular, adaptive systems that outperform conventional approaches in both user satisfaction and operational effectiveness.

3 System Architecture & Methodology

The proposed multi-modal guidance system integrates several selected hardware components to provide interactive, adaptive guidance in industrial scenarios as shown in Fig. 1. **Note**: Only the hand tracking mode for interaction is utilized for this work [1]. The motion tracking system with an interactive pen is included in the demonstrator but is not considered in the scope of this work.

Hardware Components:

- Camera and projector: A camera-projector system is utilized to capture real-time images for hand tracking and to project dynamic overlays directly onto workpieces. The projector and camera are positioned to ensure optimal coverage of the workspace, particularly targeting the user's hand movements and interaction area.
- Interactive GUI for additional insights. A touch-optimized GUI, accessible on tablet-like devices, complements the projections with detailed defect and rework information. The GUI allows mode selection (Novice/Expert), visualizes defect markings, and provides performance evaluations (Fig. 2). The home screen serves as the starting point, where users select their desired operational mode. Expert Mode visualizes previously assessed defects and assists the user by suggesting specific rework steps. Novice Mode helps train new users by visualizing and interacting with marked defects. It also provides a performance evaluation (Evaluation Results) by comparing the user's defect markings to ground truth data.
- Projected UI: Displays real-time system status and defect details directly on the work surface for seamless user feedback.
- Interaction device (Presenter): Enables easy navigation through system options, including defect type selection.

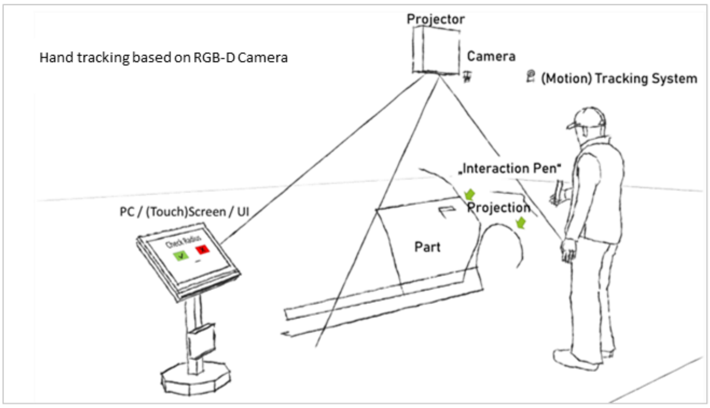

Fig. 1: Schema of the multi modal user guidance setup.

Software Framework:

- **Real-time hand tracking and gesture recognition**. Using SketchGuide [1] methods, 2D hand keypoints are extracted and back-projected into 3D space for precise, marker-free defect marking.
- **Projection mapping and dynamic overlay generation**. Calibrated projection mapping algorithms are implemented to dynamically overlay precise visual cues and guidance directly onto physical parts. Calibration between the system and the workpiece is achieved by the user touching three predefined reference points using the right-hand fingertip. Calibration accuracy via fingertip-based selection of three predefined points typically achieves sufficient precision (<5 cm). The GUI visually displays both reference and user-tipped points, providing immediate feedback to assess and confirm calibration quality. Synchronization between the projector output and user interactions ensures the overlays adapt immediately to the user's actions and expertise level, providing contextual and task-specific visual assistance.
- **Integration of 3D models for enhanced visualization**. Modern web frameworks (React/Three.js) enable cross-platform, screen-size-agnostic interaction with component part models in **real-time**. Part geometry and metadata are dynamically loaded from a backend database (Redis) upon user request, allowing immediate interaction within the browser. Marked regions are **computed on-the-fly** as users trace defect contours with their fingertip positions, considering the calibrated projector's spatial alignment. These regions, along with defect types, are stored alongside the 3D model data for future retrieval. Users can interact with these markings in real-time by selecting defect regions within the 3D view or through the side panel, triggering projections or accessing rework information with synchronized frontend-backend updates.

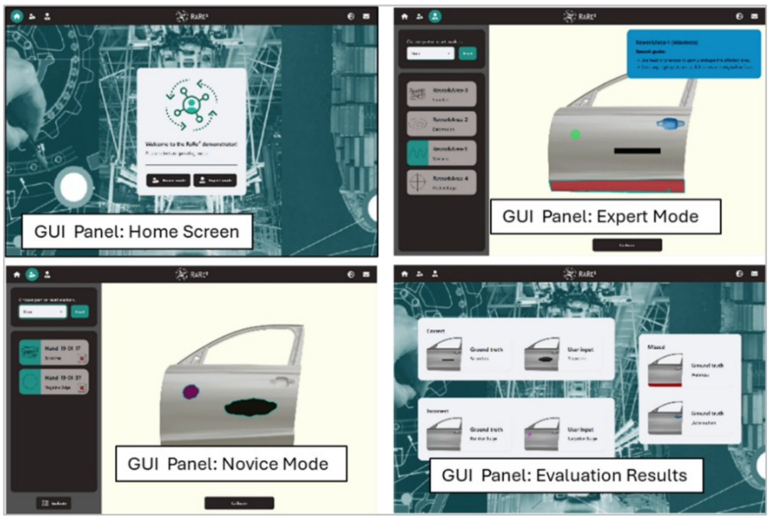

Fig. 2: Different modes of interactions on the GUI, including Novice Mode and Expert Mode.

Modes of Operation:

- **Novice Mode**: Trains new users by providing interactive guidance through projected overlays and ground truth comparisons. It allows users to visualize and interact with marked defects while a backend module evaluates their input against verified ground truth data. Performance evaluation is triggered at the end of the marking process, helping workers develop the necessary skills with greater precision and confidence.
- **Expert Mode**: Assists process experts by projecting defects directly onto the parts and displaying specific rework instructions on the GUI. This mode ensures a seamless workflow by integrating contextual information and rework hints, reducing the need for external rework guides and facilitating automated integration into product sheets.

4 Experimental Setup, Results and Discussions

4.1 Experimental Setup

As shown in Fig. 3, the setup includes an RGB-D camera (Orbbec Astra 2), a projector (Panasonic PT-MZ880), an interactive tablet (Microsoft Surface Pro 4), and a sample car door for defect visualization and rework guidance. The ceiling-mounted RGB-D camera (**1280 × 720**) tracks real-time hand and fingertip positions, with defect types selectable via a presenter. The projector, mounted above the part, overlays defect locations and color-coded labels. Operating at **1920 × 1200** resolution, it provides clear visual feedback, including fingertip positions and defect labels, ensuring accurate projection alignment across varied surfaces. The tablet serves as a control panel for mode selection, displaying defect information, rework hints, and training results. It also allows users to manage defect markings and projections with real-time feedback. The **accuracy of defect marking** using real-time hand tracking and projection feedback aligns with the evaluation results presented in **SketchGuide [1]**, which the system builds upon for precise, markerless sketching interactions. The system typically achieves sufficient accuracy (< 5cm) under calibrated conditions, sufficient for typical defect annotation tasks in industrial rework scenarios.

4.2 Results and Discussion

A detailed demonstration of the multi-modal guidance system can be accessed in this video[1]. Fig. 4 shows an excerpt of the system in action. In **Novice Mode**, the user selects the testing model via a drop-down menu and marks identifiable defect regions on the part using hand tracking (index finger position). The specified regions are projected onto the part and visualized in the GUI, where projections can be toggled on or off, and mis-markings can be deleted. Classification of defect types is managed by cycling through a list of known defects using a presenter device. In **Expert Mode**, the user is presented with a list of pre-determined defects, which can be projected onto the test part by selecting the region in the 3D view or choosing the corresponding item from the sidebar. Additionally, a defect-specific rework guideline is displayed in the top-right corner when a particular defect is selected.

The implementation and evaluation of the proposed multi-modal guidance system demonstrates its effectiveness in enhancing training and rework processes within industrial settings. By integrating real-time hand tracking, projection-based visual guidance, and interactive GUIs, the system enables intuitive, marker-less interaction, reducing reliance on traditional tools like pens or printed instructions. Its dual-mode operation offers adaptive support tailored to user skill levels—Novice Mode facilitates skill development through real-time feedback and comparison with ground truth data, while Expert Mode provides targeted visual cues and contextual rework instructions for efficient task execution. However, deployment revealed challenges such as the sensitivity of projector-camera-workpiece calibration to part repositioning, reduced projection visibility under high ambient light, and diminished hand tracking accuracy due to occlusion or glove use. To address these issues, the system includes GUI-based redundancy to maintain guidance under suboptimal conditions, and an internally tested OptiTrack-based tracking system with an interactive pen offers higher defect marking accuracy, though its complexity limits broader adoption. However, the added hardware complexity and cost present scalability concerns, which is why this configuration remains outside the scope of the current work. Overall, the system shows strong potential to enhance human-machine interaction in HMLV environments, improving engagement, reducing errors, and shortening training cycles. Its modular design supports broader deployment beyond the automotive sector.

---

[1] https://youtu.be/RN4L7yWaR6g

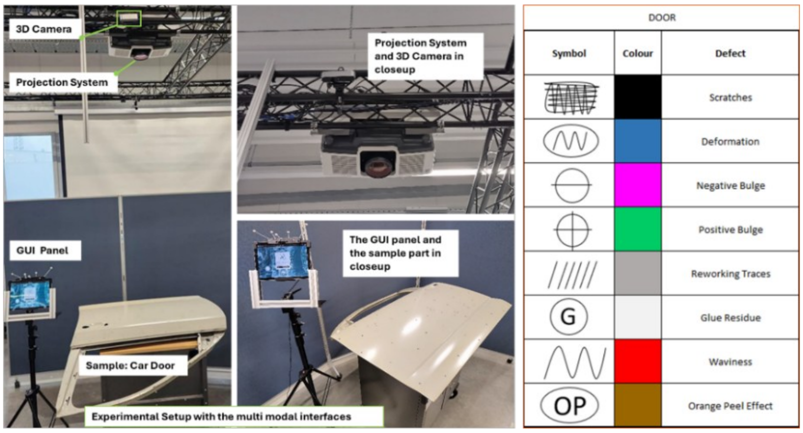

Fig. 3: Experimental setup and supported defect types and color encodings.

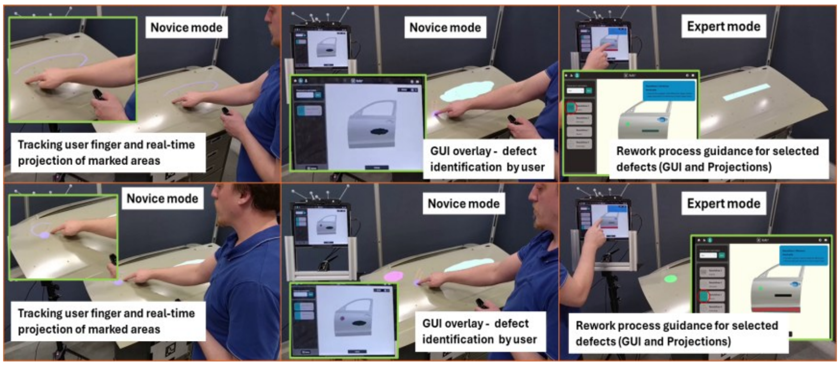

Fig. 4: The multi-modal guidance system in action: Novice Mode and Expert Mode.

5 Conclusion & Future Work

This work presented a skill-adaptive, multi-modal guidance system for improving training and rework in industrial environments. By integrating real-time hand tracking, projection-based feedback, and interactive GUIs, the system offers intuitive, context-aware guidance tailored to user expertise. Its dual-mode operation supports novices with interactive training and experts with targeted rework assistance, leading to higher task efficiency, reduced hardware needs, and smoother workflows. Experimental results confirmed improved task performance, user engagement, and effective reuse of parts, making the system well-suited for high-mix, low-volume production where flexibility is essential.

Despite its demonstrated effectiveness, several limitations were observed, including calibration sensitivity to part repositioning, reduced projection visibility in bright environments, and decreased hand tracking accuracy with gloves or occlusion. Future work will address these through RGB-D-based automatic calibration, glove-compatible hand tracking, and better synchronization of projections and GUI elements. Additional input methods like voice control will be explored to enhance usability. The system will be further tested in sectors such as electronics and aerospace, with ongoing development of gamification elements to support motivation and learning outcomes.

Acknowledgements

This research is funded by the EU funded project RaRe2 (Grant agreement number: 101092073) and by the project DemoDatenPro (Wi-2022-603642), an FTI Initiative Kreislaufwirtschaft funded by the Upper Austrian Government.

References

1. Heindl, C., Ebenhofer, G., Kutlu, S., Widmoser, F., Pichler, A.: SketchGuide: A Baseline Vision-Based Model for Rapid Robot Programming via Freehand Sketching on Any Surface. In: Huber, M., et al. (eds.) ERF 2025, SPAR, vol. 36. Springer, Cham (in press).
2. Seeliger, A., Cheng, L., Netland, T.: Augmented reality for industrial quality inspection: An experiment assessing task performance and human factors. *Computers in Industry* 151, 103985 (2023).
3. Havlíková, K., Hořejší, P., Kopeček, P.: Effect of augmented reality support on quality inspection of welded structures. *Applied Sciences* 13(21), 11655 (2023).
4. Suzuki, R., Karim, A., Xia, T., Hedayati, H., Marquardt, N.: Augmented reality and robotics: A survey and taxonomy for AR-enhanced human-robot interaction and robotic interfaces. In: *Proc. CHI 2022*, pp. 1–33. ACM, New York (2022).
5. Wedral, A., Vrecar, R., Ebenhofer, G., Pönitz, T., Wührer, P.H., Weiss, A., Stübl, G.: Spatial Augmented Reality in the Factory: Can In-Situ Projections Be Used to Communicate Dangers and Health Risks?. In: IFIP Conference on Human-Computer Interaction, pp. 574–594. Springer, Cham (2023).
6. Puthenkalam, J., Zafari, S., Sackl, A., Gallhuber, K., Ebenhofer, G., Ikeda, M., Tscheligi, M.: Where should I put my Mark? VR-based Evaluation of HRI Modalities for Industrial Assistance Systems for Spot Repair. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 2148–2155. IEEE, (2023).
7. Pönitz, T., Ebenhofer, G., Stübl, G., Heindl, C., Scharinger, J.: On the potential of large-scale extended reality interaction for industrial environments. In: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 61–63. ACM, New York (2021).
8. Stübl, G., Ebenhofer, G., Bauer, H., Pichler, A.: Lessons learned from industrial augmented reality applications. *Procedia Computer Science* 200, 1218–1226 (2022).
9. Stübl, G., Heindl, C., Ebenhofer, G., Bauer, H., Pichler, A.: Lessons learned from human pose interaction in an industrial spatial augmented reality application. *Procedia Computer Science* 217, 912–917 (2023).
10. Ikeda, M., Widmoser, F., Stübl, G., Zafari, S., Sackl, A., Pichler, A.: An Interactive Workplace for Improving Human Robot Collaboration: Sketch Workpiece Interface for Fast Teaching (SWIFT). In: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2023 ACM International Symposium on Wearable Computing, pp. 191–194. ACM, New York (2023).