LaserCoPilot: A Transferable Multimodal Co-pilot for Closed-Loop Installation and Assembly Guidance via Step-Indexed Laser Projection

Mengjun Wang[1], Ruiren Mei[2], Fan Xu[3] and Shuai Li[4*]

[1]Department of Civil & Coastal Engineering, University of Florida, United States
[2]Department of Mechanical & Aerospace Engineering, University of Florida, United States
[3]Department of Electrical & Computer Engineering, University of Florida, United States
[4]Department of Civil & Coastal Engineering, University of Florida, United States (Corresponding Author*)

mwang5@ufl.edu, ruirenmei@ufl.edu, xu.fan@ufl.edu, shuai.li@ufl.edu

Abstract -

Manual assembly in construction and manufacturing currently relies on static instructions (manuals, SOPs) that are brittle to variability and lack real-time feedback. To address this, we present a laser-based multimodal co-pilot that delivers dynamic, closed-loop guidance in non-wearable settings. The proposed framework compiles heterogeneous documentation (manuals, CAD/BIM) into a unified representation containing a step graph and verifiable criteria. At runtime, the system fuses user intent, scene perception, and uncertainty modeling to generate synchronized natural-language guidance and actionable laser cues. Crucially, it performs post-action verification to assess quality and issue targeted corrections, while maintaining persistent session memory for automated reporting. Proof-of-concept validation demonstrates the system's feasibility for interactive guidance, grounded Q&A, and robust quality verification.

Keywords -

Spatial augmented reality; work instructions; vision-language models; progress tracking; quality verification

1 Introduction

Across construction, industrialized building manufacturing (IBM), and factory-floor manual assembly, frontline work is still dominated by paper manuals, 2D drawings, and informal know-how, rather than context-aware, in-situ guidance [1, 2, 3]. This reliance becomes especially brittle under high product variability, frequent design revisions, and heterogeneous worker experience, where ambiguity about "which part/which orientation/which hole/which sequence" readily leads to mis-assembly, omissions, and costly rework [4, 3]. In construction, a persistent productivity decline over the last five decades has been documented: value added per worker in 2020 was reported to be roughly 40% lower than in 1970, even as aggregate U.S. productivity increased substantially [5]. Rework is a major contributor, and industry surveys estimate that in the U.S. alone approximately $65B was spent on rework in 2018, with nearly half attributed to inaccurate, inaccessible, or incompatible project information (i.e., information dislocation/ambiguity) [6]. Similar "information-to-action" gaps arise in manufacturing and IBM: manual assembly performance and error risk are strongly influenced by task complexity and the quality and format of work instructions, and increased cognitive/physical demands are associated with higher human error likelihood [4, 7, 8]. Moreover, despite known drawbacks, paper-based drawings remain widely used in practice, and their cross-referencing burden and ambiguity can contribute to misinterpretation and downstream errors [2].

To address these challenges, digital work instructions (DWI) and augmented reality (AR) interfaces have been explored for step-wise assistance. Prior studies show that AR-based instruction delivery can improve performance compared to traditional manuals in assembly/disassembly tasks [1], yet real deployments frequently face barriers such as wearable burden, calibration/maintenance overhead, and the cost of authoring and updating AR content at scale [9, 10, 7]. These limitations have motivated increasing interest in spatial augmented reality (SAR), which projects guidance directly onto the workspace and keeps workers hands-free while providing a shared visual reference. Industrial systems demonstrate practical feasibility of projection-based "virtual templating" and CAD-driven laser guidance for positioning and assembly [11, 12]. Construction research has likewise explored BIM-/model-driven projection paradigms [13, 14, 15]. However, most existing DWI/AR/SAR systems remain fundamentally *static*: they can display pre-authored steps or geometric cues, but they struggle to (i) interpret the current state of an unstructured workspace, (ii) infer progress when the worker deviates from the nominal plan, and (iii) provide on-demand clarification and correction when ambiguity arises.

This paper closes the loop by introducing a laser-based multimodal co-pilot for non-wearable, workstation-centric

guidance in installation and assembly tasks. The core idea is to combine step-indexed laser projection (for precise, actionable visual cues) with vision-language reasoning (for scene-aware progress inference and natural-language clarification), enabling dynamic, interactive guidance rather than static overlays. This co-pilot system is designed to be hardware-agnostic, where workers request assistance via voice/gesture/button, the system interprets the workspace state from visual evidence, and guidance is delivered through synchronized audio and laser-projected cues. Concretely, we make three contributions:

1. **A transferable co-pilot framework** that compiles heterogeneous task documentation (manuals and/or CAD/BIM artifacts) into a unified intermediate representation (IR) with step graphs, visually verifiable completion criteria, and a step-indexed laser pattern database.
2. **A closed-loop interaction loop** that fuses user intent, scene evidence, and compiled task knowledge to infer task progress, generate synchronized verbal and laser guidance, and perform post-action verification with step-level quality scoring and targeted auto-correction.
3. **A proof-of-concept validation** demonstrating the feasibility of interactive guidance for unstructured workbench assembly, highlighting how the co-pilot supports both step execution and on-demand Q&A grounded in the current workspace context.

2 Related Studies

In traditional construction settings, AR technologies have been widely studied and adopted to enhance worker efficiency and reduce errors. AR has been employed for worker training and real-time visualization of construction blueprints and BIM content [16, 17], helping workers interpret complex geometry and sequencing. Applications also include overlaying digital instructions on physical objects for task execution, as well as facilitating remote collaboration through AR interfaces where experts provide real-time annotations and step guidance [18, 19]. Similar needs extend beyond construction to manufacturing and IBM, where frontline operators often rely on paper-based SOPs, 2D drawings, or tablet-based instructions that are costly to maintain and prone to interpretation errors under high variability. Recent studies in manufacturing continue to report that instruction format and presentation significantly affect assembly time, error rates, and operator workload, motivating more situated and visually grounded instruction delivery [4]. Despite these advantages, AR systems still face notable limitations in real deployments: wearable devices can be bulky and impose cognitive/physical burden; handheld AR interrupts hands-free work; multi-user sharing is limited; and registration precision is often insufficient for tolerance-sensitive tasks and fine-grained layout guidance [9, 13].

To address these challenges, researchers have increasingly focused on SAR, which enables multiple users to interact with projected content in shared physical spaces without head-mounted displays. SAR technologies are typically implemented using either digital projection or laser projection systems. Digital projectors have been used to project BIM data onto work surfaces and surrounding geometry, guiding workers through installation and layout processes [20, 13]. In parallel, projection-based instruction systems in manufacturing have advanced from display-only guidance toward tighter integration with registration and quality inspection, for example, deep-learning-assisted projection/AR work instruction pipelines that support both guidance and inspection in real time [21]. Laser-based SAR systems offer highly precise, high-resolution projections on surfaces, enabling detailed CAD model visualization for structural layout marking, welding-point identification, and precision cutting [22, 23]. More recent work continues to strengthen the laser SAR direction by formalizing end-to-end pipelines for generating projection imagery from geometric primitives (e.g., line segments) and validating compact camera–laser configurations in real outdoor/industrial conditions [24]. Moreover, emerging BIM-to-laser workflows emphasize hardware-agnostic compilation from design files to projection instructions, supporting greater transferability across systems and deployment contexts [15]. Collectively, these studies demonstrate that SAR can reduce reliance on wearable devices while improving the precision and shareability of visual guidance across construction and manufacturing settings.

Despite the progress of AR/SAR, a key limitation persists: most systems remain largely *static* and offer limited *adaptive interactivity*. Construction sites and assembly stations are inherently dynamic, workers deviate from nominal plans, encounter ambiguity (e.g., part identification, orientation, sequence), and require on-demand clarification. Mobile SAR solutions [25, 14] expand where projection can occur, yet often emphasize mobility and projection placement rather than interactive knowledge exchange and closed-loop reasoning. Recent manufacturing-oriented SAR systems have begun incorporating stronger closed-loop behaviors, for instance, gesture-controlled SAR that supports adaptive sequencing and real-time error detection compared against paper/tablet instructions [26], but generalizable, knowledge-grounded assistants that can both *explain* and *verify* remain limited. This gap has catalyzed interest in "co-pilot" paradigms that combine situational awareness with interactive reasoning. In other industries, co-pilots provide context-aware support and

adaptive decision assistance [27, 28]. Analogously, AEC research is increasingly exploring LLM- and agent-based assistants that enable natural-language access to project knowledge (e.g., BIM querying and conversational interaction) and can support field personnel through explanation and information retrieval [29, 30, 31]. These trends point to an emerging direction: unifying projection-based, step-indexed visual cues with LLM/VLM-enabled scene-aware reasoning to enable closed-loop guidance that performs progress inference, on-demand clarification, verification, and reporting.

Overall, the literature highlights the need for a *non-wearable, workstation-ready co-pilot* that moves beyond static overlays by integrating (i) design/manual-derived step knowledge and projection assets, (ii) scene-aware progress inference, and (iii) closed-loop verification with interactive, real-time clarification supporting unstructured installation and assembly tasks across construction, IBM, and manufacturing.

3 Methodology

We present a laser-based multimodal co-pilot that provides dynamic, closed-loop guidance for human-performed installation and assembly tasks. The system integrates (i) worker-initiated interaction (voice, gesture, or button), (ii) scene-aware reasoning via vision–language models, and (iii) step-indexed laser projection coupled with natural-language explanations. The co-pilot continuously estimates task progress, delivers synchronized audio and visual instructions, verifies outcomes, produces quality scores and progress reports, and issues corrective feedback when deviations are detected.

To support deployment across diverse tasks, the framework is designed to be transferable across installation manuals and design sources. Transferability is achieved by compiling heterogeneous documentation (installation manuals, CAD/BIM models, and design drawings) into a unified IR and a laser pattern database. At runtime, the co-pilot reasons over the compiled task representation and current scene evidence to determine what to project and what to communicate, while maintaining persistent session memory to stabilize step advancement and summarize performance. Figure 1 describes the high-level system components and functions.

3.1 System framework

The co-pilot is organized as a modular pipeline comprising four components: *Knowledge*, *Perception*, *Decision*, and *Verification & Memory*. Offline, the system compiles heterogeneous task sources into a transferable representation, including a step graph, visually verifiable completion criteria, and a step-indexed laser pattern database (built on

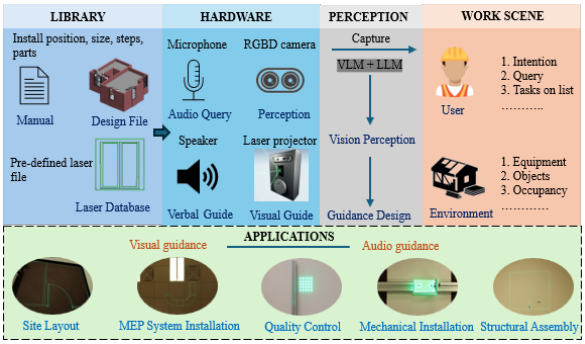

Figure 1. System overview

our prior laser file generation pipeline). During operation, a worker triggers assistance, after which the co-pilot captures the workspace and estimates a structured scene state with explicit uncertainty modeling. The decision module then fuses the scene state, user intent, and compiled task knowledge to infer progress, select an appropriate guidance strategy, and generate synchronized verbal responses together with laser projections resolved deterministically from the database. After the worker executes the instruction, the co-pilot performs closed-loop verification to evaluate step completion, assign step-level quality scores, and issue targeted corrective actions when deviations are detected. Finally, session memory aggregates step status, evidence, and scores to stabilize progress tracking and enable automated reporting. Figure 2 illustrates the overall workflow.

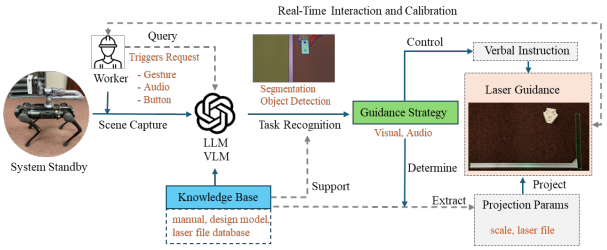

Figure 2. The proposed framework workflow

3.2 System architecture

This section details each system module and illustrates the connection in Figure 3, which summarizes the closed-loop interaction procedure.

3.2.1 Knowledge layer

The co-pilot relies on an offline *knowledge compilation* stage that converts heterogeneous task documentation (installation manuals, CAD/BIM models, and design drawings) into a transferable task package. We represent each

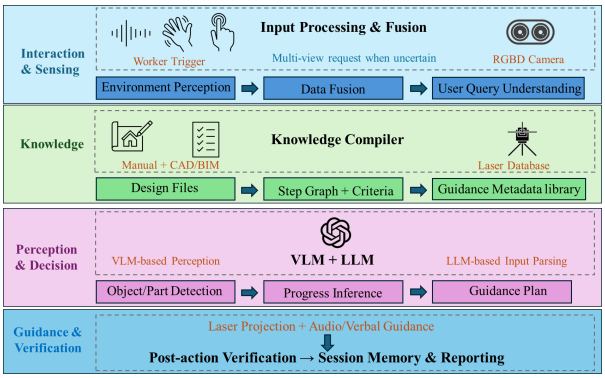

Figure 3. The system architecture

task as a directed step graph $G = (V, E)$, where nodes correspond to steps (or micro-steps) and edges encode precedence, optional branches, and rework transitions. Each step is augmented with a compact set of *visually verifiable completion criteria* that can be grounded in perceptual evidence (e.g., part presence, attachment status, fastener evidence, and alignment within tolerance).

For visual guidance, each step is linked to one or more laser patterns stored in a step-indexed laser database $\mathcal{L}$ with associated projection metadata. The laser projection hardware and the conversion pipeline from design artifacts to laser projection files are built on our prior work [15]. In this study, we focus on the co-pilot's reasoning and closed-loop guidance mechanism; at runtime, projected patterns are resolved deterministically from $\mathcal{L}$ rather than generated ad hoc, ensuring auditable and non-hallucinatory actuation.

3.2.2 Perception layer

Upon worker initiation (voice, gesture, or button), the system captures the current workspace using one or more cameras and estimates a structured scene state s. The scene state summarizes task-relevant evidence, including component/tool presence, spatial relations and attachment hypotheses (e.g., positioned/attached/tightened), fastener evidence (e.g., visibility and approximate placement), and observation quality indicators (occlusion/blur/glare). The perception module is uncertainty-aware: when evidence is insufficient or occluded, attributes are marked *unknown* rather than extrapolated, enabling conservative downstream reasoning and targeted clarification requests.

3.2.3 Decision layer

The decision layer fuses the worker intent q, the estimated scene state s, and the compiled task package $\mathcal{T} = (G, \mathcal{K}, \mathcal{L})$ to infer progress and generate actionable guidance. First, the worker input is parsed into an intent class (e.g., *next step*, *verify*, *how-to question*, *recover*). Second, the system infers the most likely current step by matching s against candidate steps and their completion criteria, optionally constrained by session history to enforce stable, monotonic advancement unless rework is requested.

Based on the inferred state, the co-pilot selects a guidance strategy consisting of (i) the target step/sub-step, (ii) the corresponding laser pattern(s) resolved from $\mathcal{L}$, and (iii) a verbal response that provides concise instructions, disambiguation, and self-check cues. When confidence is low, the co-pilot requests additional evidence (e.g., alternate viewpoints or close-ups) or poses targeted confirmation questions, rather than issuing potentially incorrect instructions.

3.2.4 Verification and memory

After the worker executes the instruction, the system captures post-action observations to verify whether the target step's completion criteria are satisfied. Verification produces a pass/fail decision, a step-level quality score (e.g., 0–100), and localized deviation descriptions. If deviations are detected, the co-pilot issues corrective guidance (and, when appropriate, error-localizing laser cues) and repeats verification until the step is completed or the user overrides.

All interactions are logged in a persistent session memory $\mathcal{M}$ that stores step status, evidence snapshots, quality scores, and correction events. This memory stabilizes progress tracking, supports controlled rework transitions, and enables automated reporting (e.g., step-wise progress, quality distributions, and common error patterns) for documentation and supervision.

4 Experimental Validation

Case study and task materials. To qualitatively validate the interactivity of the proposed co-pilot, we selected a workbench assembly task based on a metal frame installation manual. The manual specifies a step-wise procedure with explicit part lists, instructions, checks, and cautions for each step, following a laser-alignment convention (i.e., aligning the outer profile of parts to the projected top-view outline before tightening fasteners). Concretely, our case study manual contains 14 steps (Steps 1–14), each describing required components and step-specific constraints (e.g., do-not-tighten-before-alignment and safe tightening order). For example, several steps explicitly require matching the bracket/plate outer profile to the laser outline and alternating tightening to prevent shifting.

Manual-to-knowledge compilation. We instantiated the knowledge layer by converting the manual into a uni-

fied, machine-readable IR in JSON, which is used as the agent's knowledge library. The IR is generated by an automated compilation pipeline that parses the manual and outputs a structured `steps_ir.json` under a predefined schema, including (i) step definitions, (ii) visually verifiable completion criteria, and (iii) a step-indexed `laser_key` that binds each step to its corresponding laser pattern. The compilation imposes constraints that each step must include multiple visually checkable criteria and that `laser_key` equals `step_id` to ensure deterministic laser-file resolution during runtime. In addition, we allow lightweight manual review and refinement of the generated IR (e.g., correcting ambiguities in completion criteria) to reduce systematic bias and improve step discriminability, consistent with the guidance in our project logic documentation. Figure 4 illustrates an excerpt of the extracted components (type and quantity), the corresponding step-level design drawing, and an on-site photo of laser-projected guidance.

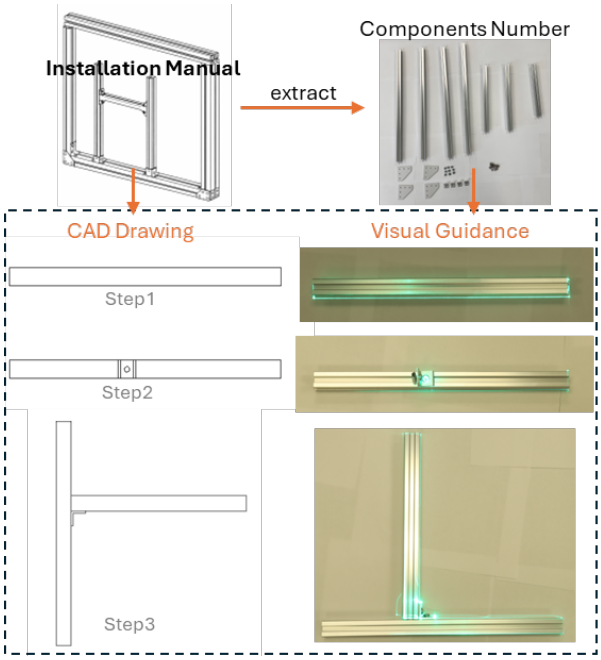

Figure 4. An installation steps example

Runtime model and simplified interaction interface. We used the GPT-5.2 API as both the vision-language module for scene understanding and the reasoning module for decision-making. To simplify the experimental workflow while preserving the core concept, we implemented a GUI that replaces continuous video/audio streams with *image upload + text input*. This interface is functionally equivalent to the full multimodal setting, as voice/gesture/button triggers can be integrated as alternative front-ends without altering the co-pilot's reasoning loop.

System interface. The interface contains seven primary components that operationalize closed-loop guidance (See Figure 5): (1) *User options* that specify the current intent mode (e.g., scene reference, correctness check, or open-ended Q&A); (2) a *scene upload panel* supporting 1–3 images per query, consistent with our perception pipeline; (3) an *inference panel* reporting the estimated current step, confidence, step-specific instructions, cautions, and the laser file to call, while reminding the user to perform post-action verification; (4) a *user feedback panel* for rating the agent's guidance, which updates session memory; (5) a *free-form query panel* that supports natural language questions grounded in the current task context (e.g., identifying a small part by appearance and explaining its usage in subsequent steps); (6) a *verification panel* that accepts post-action images and outputs a step-level performance score (0–100) together with targeted improvement suggestions, repeating verification if necessary; and (7) a *side panel* that summarizes progress (current/next step, completed steps) and aggregates performance statistics for documentation.

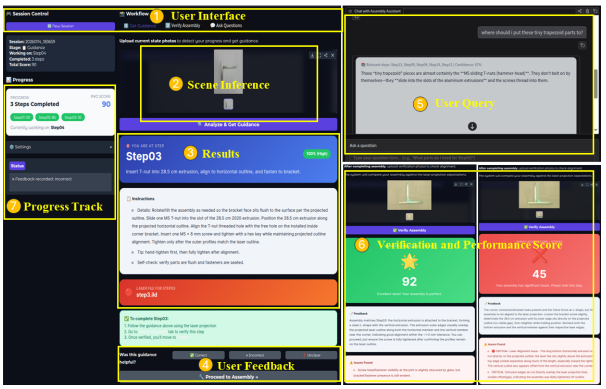

Figure 5. The co-pilot interface

Closed-loop procedure and qualitative findings. During each interaction round, the user uploads the current workspace image(s), the agent constructs a structured scene state (`SceneState`) via the vision API, retrieves top-k candidate steps via RAG, and scores candidates by completion-criteria satisfaction to infer the most likely current step and the next guidance action. The workflow then prompts the user to upload a post-action image for verification; the verification module assigns a step-level score, logs evidence and outcomes into session memory, and advances the step only when performance exceeds a minimum acceptance threshold. Across the case study, this experiment qualitatively demonstrates the feasibility of our framework as an interactive, closed-loop co-pilot

for unstructured assembly guidance, supporting both multimodal instruction (audio and laser) and iterative correction with progress and quality tracking.

5 Conclusion

This paper introduced a transferable, laser-based multimodal co-pilot for dynamic guidance in installation and assembly tasks. Unlike traditional digital work instructions and many AR/SAR systems that primarily present static, pre-authored overlays, the proposed framework closes the loop by integrating (i) offline compilation of heterogeneous task sources into a step-structured intermediate representation with step-indexed laser patterns and verifiable completion criteria, (ii) scene-aware progress inference and on-demand explanation via vision-language reasoning, and (iii) post-action verification with step-level quality scoring, targeted correction, and persistent session memory for progress logging and report generation. By coupling precise, actionable laser-projected cues with natural-language clarification grounded in both task knowledge and observed workspace state, the co-pilot supports hands-free, shared-reference guidance without requiring wearable displays.

The proof-of-concept validation indicates that the framework can (1) infer the current step from workspace observations, (2) select appropriate laser cues and generate step-specific verbal guidance, (3) answer free-form worker questions grounded in the ongoing task context, and (4) verify outcomes to support reliable step advancement and quality documentation. These results suggest that knowledge-compiled, scene-aware co-pilots offer a promising direction for reducing ambiguity-driven errors and rework across construction, IBM, and general manual assembly.

Future work will extend the framework in three directions. First, we will strengthen transfer by expanding the intermediate representation to support multi-branch step graphs, tool/part ontologies, and robust version control across design revisions. Second, we will improve perception and verification under occlusions and diverse viewpoints via multi-frame evidence aggregation and uncertainty-aware decision policies. Third, we will integrate real-time multimodal interaction (live video, speech, and gesture/button triggers) and evaluate the system at larger scale with quantitative measures of time, error rate, cognitive load, and rework reduction in representative industrial settings.

Acknowledgement

This research was funded by the National Science Foundation (NSF) via Grant 2222810 and Grant 2524239. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF and The University of Florida.

References

[1] H. Dorloh, K.-W. Li, and S. Khaday. Presenting job instructions using an augmented reality device, a printed manual, and a video display for assembly and disassembly tasks: What are the differences? *Applied Sciences*, 13(4):2186, 2023. doi:10.3390/app13042186.

[2] M. Foroughi Sabzevar, M. Gheisari, and J. Lo. Analyzing the pros and cons of paper-based 2d drawings in construction: A survey of u.s. construction professionals. *International Journal of Construction Management*, 25(4):428–438, 2025. doi:10.1080/15623599.2024.2331863.

[3] M. Wang, J. Cai, D. Hu, Y. Hu, Z. Han, and S. Li. AI-based robots in industrialized building manufacturing. *Frontiers of Engineering Management*, 12 (1):59–85, 2025. doi:10.1007/s42524-025-4099-x.

[4] M. A. Noman, F. M. Alqahtani, W. Ameen, M. H. Alhaag, A. A. Ahmed, and A. E. Abdelgawad. Impact of instruction methods and task size on human performance and physical workload in manual assembly. *IEEE Access*, 2025. doi:10.1109/ACCESS.2025.3637395.

[5] A. Goolsbee and C. Syverson. The strange and awful path of productivity in the US construction sector. On-line: `https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us` Accessed: 2026-01-15.

[6] PlanGrid and FMI. Construction disconnected 2018 industry report. On-line (PDF): `https://pg.plangrid.com/rs/572-JSV-775/images/Construction_Disconnected.pdf`. Accessed: 2026-01-15.

[7] L. Bock, T. Bohné, and S. K. Tadeja. Decision support for augmented reality-based assistance systems deployment in industrial settings. *Multimedia Tools and Applications*, 84(21):23617–23641, 2025. doi:10.1007/s11042-024-19861-x.

[8] B. Klages, A. Metzner, and M. F. Zaeh. A methodology for the prioritization of factors influencing human errors in manual assembly lines. *Production Engineering: Research and Development*, 19:1341–1358, 2025. doi:10.1007/s11740-025-01370-x.

[9] N. Ashtari, A. Bunt, J. McGrenere, M. Nebeling, and P. K. Chilana. Creating augmented and virtual reality applications: Current practices, challenges, and opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, pages 1–13. Association for Computing Machinery, 2020. doi:10.1145/3313831.3376722.

[10] Y. Yin, P. Zheng, C. Li, and L. Wang. A state-of-the-art survey on augmented reality-assisted digital twin for futuristic human-centric industry transformation. *Robotics and Computer-Integrated Manufacturing*, 81:102515, 2023. doi:10.1016/j.rcim.2022.102515.

[11] FARO. FARO tracer laser projectors (product/resource pages). On-line: `https://www.faro.com/en/Products/Hardware/Tracer-Laser-Projectors-for-Construction`. Accessed: 2026-01-15.

[12] LAP GmbH Laser Applikationen. CAD-PRO laser projectors. On-line: `https://www.lap-laser.com/products/cad-pro/`. Accessed: 2026-01-15.

[13] A. Degani, W. B. Li, R. Sacks, and L. Ma. An automated system for projection of interior construction layouts. *IEEE Transactions on Automation Science and Engineering*, 16(4):1825–1835, 2019. doi:10.1109/TASE.2019.2897135.

[14] Kuo-Chiang Yeh, Ming-Hsiang Tsai, and Shih-Chung Kang. On-site building information retrieval by using projection-based augmented reality. *Journal of Computing in Civil Engineering*, 26 (3):342–355, 2012. doi:10.1061/(ASCE)CP.1943-5487.0000156.

[15] M. Wang, J. Xu, H. A. Lassiter, and S. Li. BIM-driven laser spatial augmented reality for in-situ layout and assembly. *Automation in Construction*, 178: 106405, 2025. doi:10.1016/j.autcon.2025.106405.

[16] Jad Chalhoub and Steven K. Ayer. Exploring the performance of an augmented reality application for construction layout tasks. *Multimedia Tools and Applications*, 78:35075–35098, 2019. doi:10.1007/s11042-019-08063-5.

[17] M. K. Rohil and Y. Ashok. Visualization of urban development 3d layout plans with augmented reality. *Results in Engineering*, 14:100447, 2022. doi:10.1016/j.rineng.2022.100447.

[18] F. Liu and S. Seipel. Precision study on augmented reality-based visual guidance for facility management tasks. *Automation in Construction*, 90:79–90, 2018. doi:10.1016/j.autcon.2018.02.020.

[19] B. Marques, S. Silva, J. Alves, A. Rocha, P. Dias, and B. S. Santos. Remote collaboration in maintenance contexts using augmented reality: insights from a participatory process. *International Journal on Interactive Design and Manufacturing*, 16:419–438, 2022. doi:10.1007/s12008-021-00798-6.

[20] J. H. Bae and SangHyun Han. Vision-based inspection approach using a projector-camera system for off-site quality control in modular construction: Experimental investigation on operational conditions. *Journal of Computing in Civil Engineering*, 35 (5):04021012, 2021. doi:10.1061/(ASCE)CP.1943-5487.0000978.

[21] W. Li, A. Xu, M. Wei, W. Zuo, and R. Li. Deep learning-based augmented reality work instruction assistance system for complex manual assembly. *Journal of Manufacturing Systems*, 2024. doi:10.1016/j.jmsy.2024.02.009.

[22] P. Tavares, C. M. Costa, Luís Rocha, P. Malaca, P. Costa, A. P. Moreira, A. Sousa, and G. Veiga. Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality. *Automation in Construction*, 106:102825, 2019. doi:10.1016/j.autcon.2019.04.020.

[23] E. M. Villanueva, P. Martinez, and R. Ahmad. Target-path planning and manufacturability check for robotic CLT machining operations from BIM information. *Automation in Construction*, 158:105191, 2024. doi:10.1016/j.autcon.2023.105191.

[24] Maximilian Tschulik, Thomas Kernbauer, Philipp Fleck, and Clemens Arth. Spatial augmented reality for heavy machinery using laser projections. *Computers & Graphics*, 126:104161, 2025. doi:10.1016/j.cag.2024.104161.

[25] S. Xiang, R. Wang, and C. Feng. Mobile projective augmented reality for collaborative robots in construction. *Automation in Construction*, 127:103704, 2021. doi:10.1016/j.autcon.2021.103704.

[26] Naimul Hasan and Bugra Alkan. Gest-sar: A gesture-controlled spatial AR system for assembly assistance with adaptive sequencing and real-time error detection. *Machines*, 13(8):658, 2025. doi:10.3390/machines13080658.

[27] N. Puca and G. Guglieri. Enabling civil single-pilot operations: A state-of-the-art review. *Aerotecnica Missili & Spazio*, 2024. doi:10.1007/s42496-024-00223-7.

[28] S. Wandelt, C. Zheng, S. Wang, Y. Liu, and X. Sun. Large language models for intelligent transportation: A review of the state of the art and challenges. *Applied Sciences*, 14:7455, 2024. doi:10.3390/app14177455.

[29] D. Fernandes, S. Garg, M. Nikkel, and G. Guven. A GPT-powered assistant for real-time interaction with building information models. *Buildings*, 14:2499, 2024. doi:10.3390/buildings14082499.

[30] B. Liu and H. Chen. BIMCoder: A comprehensive large language model fusion framework for natural language-based BIM information retrieval. *Applied Sciences*, 15(14):7647, 2025. doi:10.3390/app15147647.

[31] P. Guo, H. Xue, J. Ma, and J. C. P. Cheng. Advancing BIM information retrieval with an LLM-based query-domain-specific language and library code function alignment system. *Automation in Construction*, 178:106374, 2025. doi:10.1016/j.autcon.2025.106374.