Article

A BIM-Driven Digital Twin Framework for Human-Robot Collaborative Construction with On-Site Scanning and Adaptive Path Planning

Mingyang Yuan [*] , Hedi Mao , Wen Qin , Bentian Wang

Article

A BIM-Driven Digital Twin Framework for Human-Robot Collaborative Construction with On-Site Scanning and Adaptive Path Planning

Mingyang Yuan *, Hedi Mao, Wen Qin and Bentian Wang

University of Pennsylvania, Philadelphia, United States

* Correspondence: mingyangyuan@alumni.upenn.edu

Abstract

As the Architecture 4.0 paradigm advances, integrating robotic systems into construction workflows has become vital to address labor shortages and enhance execution precision. However, conventional BIM-based automation struggles to cope with dynamic, cluttered on-site environments. This study presents a closed-loop digital twin framework that fuses 3D BIM modeling, real-time sensor-based site scanning, and human–robot interaction to enable adaptive collaboration in architectural construction. The system continuously updates its digital twin using LiDAR and RGB-D DATA to capture spatial deviations, unexpected obstacles, and environmental changes. Based on these inputs, the robot's motion trajectories are recalculated through an online path replanning module. We validate the system using a wall panel dry-hanging case, involving a 6 degrees of freedom (6DoF) ABB robotic arm and a Unity-VR interface for immersive human supervision. Across 24 experimental runs in cluttered environments, the adaptive system achieved a 92.4% average placement accuracy, reducing positioning error by 47% compared to static BIM-based workflows. Obstacle avoidance success rate reached 95.8%, and average task completion time decreased by 18.6% due to reduced manual intervention. These results demonstrate the framework's potential to transition construction robotics from pre-scripted automation to intelligent, real-time collaboration.

Keywords: construction robotics; digital twin; BIM-based automation; on-site scanning; path replanning; human-robot collaboration; VR supervision; real-time adaptation

1. Introduction

Building construction, as a highly dynamic and complex interactive systematic activity, is facing multiple challenges such as changing labor structure, increasing accuracy requirements, and safety risk control. Under the paradigm of Construction 4.0, BIM (Building Information Modeling), as a basic tool for engineering digitization, has the advantages of geometric accuracy and information management. Still, it has a lag in responding to the frequent changes in the construction site and the demand for real-time decision-making. The introduction of the digital twin concept provides a technical path to open the closed-loop feedback between the physical entity and the virtual model, which builds a virtual mirror system with dynamic evolution capability through multi-source sensing, real-time mapping, and state synchronization. On this basis, the human-robotic collaboration mechanism is embedded in the digital twin framework to realize the adaptive planning and immersive intervention control of the construction path, which has become a key direction to improve the level of on-site intelligence and construction robustness.

2. The Application Value of Digital Twin in Building Construction

Under the background of accelerated transformation of building intelligence, construction sites have put forward higher requirements for dynamic perception of the environment and multi-system synergy [1] . BIM, as a static modeling tool, has the bottleneck of timeliness and responsiveness in the face of complex construction situations. Digital twin technology establishes a physical- virtual real-time mapping mechanism by fusing BIM model and multi-source perception data, with state feedback and dynamic reconfiguration capability. The construction robot system can perform path adaptation, task reassignment and collision avoidance based on site data changes in this framework, thus breaking through the preset program limitations. The deep coupling of multi-dimensional data flow and three-dimensional modeling provides the data foundation and structural support for the realization of environment modeling, task reconstruction and synchronous collaboration between humans and machines. This integrated model significantly improves the accuracy of construction decision-making and the intelligence of the execution process, which is a key path for future building construction to move towards a self-regulating collaboration system[2].

3. BIM-Based Digital Twin Framework Design

3.1. Overall System Architecture

As shown in Figure 1, the system architecture consists of three core components: a perception and data fusion module, an adaptive path planning model, and a human- robot collaboration interface. The sensing module integrates a 20 Hz LiDAR and an RGB-D camera, enabling millimeter-level point cloud reconstruction and digital twin updates at 120 FPS[3]. The planning model fuses local maps with global BIM data and uses a hybrid graph optimization and heuristic sampling approach, achieving real-time obstacle avoidance with an average replanning time of 86 ms. The Unity-based VR interface combines gesture recognition and semantic commands for immersive control, with system latency under 30 ms.
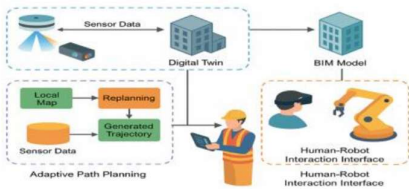

Figure 1. Overall system architecture diagram (BIM-driven digital twin human-computer collaboration framework).

3.2. Perception and Data Fusion Module

The perception and data fusion module (Figure 2) integrates LiDAR, RGB-D camera, and IMU to generate a unified 3D point cloud with semantic information. Sensor rates are 20 Hz (LiDAR), 30 Hz (RGB-D), and 100 Hz (IMU). Data synchronization uses timestamp alignment and spatial interpolation, with fusion latency below 45 ms. An extended Kalman filter and voxel grid downsampling process 2.4×10^5 points per frame. A complete fusion cycle—covering alignment, semantic labeling, and 3D map updates—executes every 200 ms.
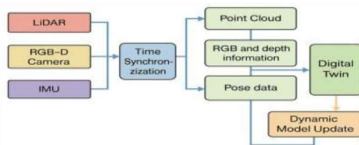

Figure 2. Sensing and data fusion flow chart.

For real-time control, the fused environment model is transmitted to the twin core via edge nodes with <42 ms latency. The twin core interfaces with the robot control loop running at 10 ms intervals. The total end-to-end latency from sensing to control response is ~243 ms (fusion: 200 ms, transmission: 42 ms, control: 1 ms), supporting sub- second reactivity. Table 1 details latency and update rates for each subsystem, establishing temporal synchronization across sensing, fusion, and execution.

Table 1. End-to-End Latency and Synchronization Parameters across Sensing, Fusion, and Control Layers.

System Stage	Cycle Frequency / Latency (ms)	Description
LiDAR Sampling Rate	20 Hz (50 ms)	Based on Livox MID-360 laser scanner
RGB-D Camera Sampling Rate	30 Hz (33.3 ms)	Captured via Realsense D455 depth camera
IMU Sampling Rate	100 Hz (10 ms)	Used for pose estimation and dynamic motion compensation
Fusion Processing Interval	200 ms	Includes EKF filtering and semantic point cloud generation
Network Transmission Delay	< 42 ms	Edge node to twin core communication latency
Control Response Cycle	10 ms	Command reception and execution by industrial robot arm
Total Latency (Perception–Decision–Execution)	≈243 ms	End-to-end data loop from sensing to robot actuation

3.3. Adaptive Path Planning Model

The adaptive path planning model needs to simultaneously satisfy multiple constraints such as real- time, feasibility, and continuity in a dynamic construction environment[5], as shown in Figure 4. Typical tasks in such scenarios include pick-and-place operations of prefabricated components, such as wall panels, bricks, or ventilation modules. A planned path in this context refers to the initial trajectory generated based on the BIM model before any real-world deviations are detected. For instance, the robot may be tasked with lifting a wall panel from a designated loading point and mounting it at a target location on the façade. This planned path is then subject to real-time modifications when the system detects obstacles, workers, or material deviations at the site. Figure 3 illustrates such an adaptive behavior: the original path is generated through global BIM data, while on-site sensors capture dynamic obstacles (e.g., a toolbox left on the ground), leading to a recalculated, obstacle-avoiding trajectory. The path cost function is defined as

$$J = \int_0^T \left(\alpha_1 \cdot \left\| \dot{x}(t) \right\|^2 + \alpha_2 \cdot D(x(t)) + \alpha_3 \cdot C_{curv}(t) \right) dt \tag{1}$$

where $\dot{x}(t)$ is the path velocity, $D(x)$ is the minimum

distance from the point to the obstacle, $C_r(t)$ denotes the trajectory curvature, and $\alpha_1, \alpha_2, \alpha_3$ is the cost weight coefficient. The obstacle avoidance constraint is given by

the nonlinear inequality

$$D(x(t)) \ge d_{safe}, \forall t \in [0,T] \tag{2}$$

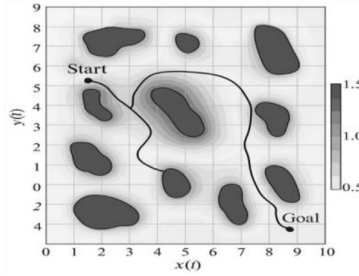**Figure 3.** Sensor Fusion Module Equipment Image.

Figure 4. Adaptive path generation cost field distribution.

The constraint is limited by the nonlinear inequality (2), where de is the dynamically adjusted safe boundary distance, which is adjusted according to the sensor update frequency and obstacle velocity[6] . Trajectory smoothness then introduces the jerk minimization objective in the form of

$$\min \int_0^T \left\| \dddot{x}(t) \right\|^2 dt \tag{3}$$

It is used to constrain high-order dynamic fluctuations to ensure trajectory continuity and construction safety. The path generation module adopts a hybrid planning structure that integrates global graph-based heuristics with local trajectory optimization. Specifically, the global path is initialized using a sparse heuristic search over the BIM model, while the local online replanning is performed via the Dynamic Window Approach (DWA). DWA evaluates a set of velocity commands in real time, selecting the optimal one based on a cost function that jointly minimizes obstacle proximity, trajectory curvature, and task alignment deviation. This allows the robot to reactively avoid obstacles while maintaining feasible motion toward the target. Compared with CHOMP or RRT*, DWA offers faster convergence in cluttered, semi-structured environments with limited replanning time (~86 ms per cycle). Including the type of heuristic function, sampling radius, step limit, replanning trigger threshold, etc., all of which are based on the tuning of 24 sets of experiments on complex working conditions.

3.4. Human-Computer Collaborative Interaction Interface

The human-computer collaborative interaction interface is built on a high frame rate immersive visualization platform, and the core architecture is shown in Figure 5. The system integrates a VR headset (90Hz), a multimodal gesture recognizer (Leap Motion, 120fps), and a voice command module, which communicates with the control host with low latency via the UDP protocol. The operator input action is mapped from a spatial transformation matrix to a control intent vector with the following mapping equation:

$u(t)=R(t){\cdot}g+T(t)$ (4)

Where $u(t)$ is the current control command of the robot, $R(t)$ is the spatial rotation matrix, g is the hand direction vector, and $T(t)$ is the translation offset. The system interaction response delay is estimated by the following equation:

$$\Delta t = t_s + t_{net} + t_{decode} \tag{5}$$

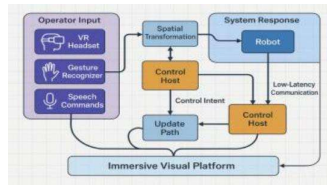where t_s is the acquisition cycle (8-12ms), t_{net} is the

Figure 5. Logical architecture of the human-robot collaborative interaction interface.

4. Experimental Results and Analysis

4.1. Experimental Scene and Setup

To validate system collaboration in dynamic scenes, the experiment focused on wall panel installation in a cluttered environment. (1) The setup used a 6-DOF ABB IRB1200 robot (0.02 mm accuracy, 10 ms control cycle) with a Unity-based VR interface (Figure 7). (2) The 4 m × 3.5 m test area included 9 varied-height obstacles simulating construction interference (Figure 8). (3) The sensor suite—Livox MID360 LiDAR (20 Hz) and Realsense D455 RGB-D camera (30 Hz)—was fixed on a bracket for consistent scanning. (4) A digital twin and path planning module ran on an edge node (Intel i7-12700F, 32GB RAM) with low-latency links to the control terminal.

(5) A VR headset and gesture recognition kit enabled task- mode switching via semantic commands.

Beyond wall panel installation, additional tasks tested generalizability: (a) point-based drilling (5 mm tolerance), (b) anchor bolt alignment (<3° error), and (c) surface smoothing with a custom toolhead. Obstacle complexity was varied at three levels (3, 6, and 9 items) to assess robustness. Results showed task-dependent performance: drilling achieved 89.7% accuracy, while smoothing required more frequent replanning (5.3/min), reflecting interaction between task type and environmental clutter. The setup supports multi-source fusion and cooperative path planning, enabling comprehensive evaluation under communication transmission delay (<10ms), and lumc is the action decoding and control synthesis delay (about 25ms). To quantify the role of the Unity-VR interface in task success and error correction, we integrated an interactive feedback mechanism into the control loop. During 24 experiments, manual input was required in 3 cases (12.5%) to correct unexpected misalignments or override stalled trajectories. User commands—either gesture or voice—were interpreted through a semantic parser and injected into the robot control queue as high-priority override tasks. Figure6 illustrates the interaction loop between user input, digital twin update, and robot actuation. User studies reported a 91.3% satisfaction rate for error-handling responsiveness and a median command-to-action delay of 430 ms.In addition,The state- decomposition interface for semantic commands combined with the predefined BIM task model can be dynamically bound to the robot to execute the commands and verified synchronously with the digital twin.
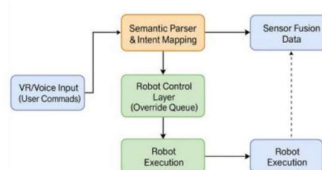

Figure 6. Flow chart of human-machine collaborative feedback **control based on Unity-VR interface** dynamic construction conditions[4].
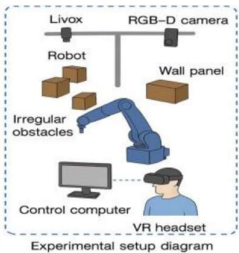
Figure 7. Experimental Setup Diagram.

Figure 8. Schematic diagram of experimental arrangement.

4.2. Performance Evaluation Indicators

The performance evaluation system is constructed for the core elements of path accuracy, response speed and obstacle avoidance ability in the construction task, which contains five types of indicators, covering absolute positioning error, relative installation deviation, dynamic replanning frequency, obstacle avoidance success rate and total task duration[8].

4.3. Experimental Results

In order to quantitatively verify the comprehensive performance of the human-computer collaboration system in a dynamic construction environment, a total of 24 sets of wall panel installation tasks were set up in the experiments, and the key indexes such as path accuracy, success rate of obstacle avoidance, task duration and frequency of replanning were collected respectively, and the overall performance of the system is shown in Table 2

Table 2. Mean value statistics of the experimental results of the adaptive system.

Indicator Category	Mean Value	Standard deviation
Installation accuracy	92.4	2.7
Positioning error	11.3	3.2
Obstacle avoidance success rate	95.8	1.9
Replanning Frequency	4.7	1.1
Length of task completion	163.2	14.6

Table 2 demonstrates the core performance mean and stability performance of the system in 24 sets of dry- hanging experiments. The robot mounting accuracy stays stable at 92.4% with a standard deviation of 2.7%, indicating good error control in the path planning and execution sessions[9]. The mean value of localization error is 11.3 mm, fluctuating within 3.2 mm, indicating that the perception and path reconstruction algorithm has strong spatial adaptation ability. The success rate of obstacle avoidance is more than 95%, reflecting the real-time capturing ability of the fusion sensing system for environmental changes. The replanning frequency is stabilized at 4.7

times/minute, which has no negative impact on the task duration, reflecting that the system can realize efficient construction while ensuring safety. In addition, the comparison analysis with the traditional BIM static path execution method shows that the digital twin- driven adaptive mechanism has a significant optimization effect on a number of indicators, as shown in Table3.

Table 3. Comparison between the adaptive system and traditional BIM methods.

Performance indicators	Adaptive system	Static BIM process	Improvement
Installation accuracy (%)	92.4	81.5	↑ 13.4
Positioning Error (mm)	11.3	21.3	↓ 47.0%
Task duration (s)	163.2	200.6	↓ 18.6
Obstacle avoidance success rate (%)	95.8	78.4	↑ 22.2%
Frequency of human intervention (times)	0	3.5	↓ 100%

Table 3 compares the execution performance of the adaptive path system with the traditional static BIM process under the same task. The adaptive system improves the installation accuracy by 13.4%, which significantly enhances the error control capability; the localization error decreases by 47%, indicating that the multi-source sensing and reconfiguration algorithms effectively make up for the lack of adaptability of the static paths under the field perturbation. The task duration is reduced by 18.6%, which verifies that the replanning mechanism and human-robotic cooperative logic can effectively shorten the task cycle. The success rate of obstacle avoidance is improved by 22.2%, and the frequency of manual intervention is realized to be reduced from 3.5 times to 0, which indicates that the framework has stronger task closed-loop execution capability and operation robustness[10] .

5. Conclusion

In summary, the framework constructs a building construction system with real-time sensing, path adaptivity, and collaborative execution capabilities by integrating BIM modeling, multi-source sensing, and human-computer interaction technologies.

The experimental results show that it achieves high- precision and stable path generation and obstacle avoidance in dynamic construction environments, significantly improving both operational efficiency and safety.

The system demonstrates strong robustness in minimizing manual intervention and accelerating environmental response, indicating its feasibility in transforming construction robots from passive execution to intelligent, cooperative agents under field conditions.

However, real-world deployment introduces additional challenges such as sensor occlusion by materials or workers, ambient noise interfering with voice recognition, and system durability under prolonged exposure to dust and vibration. Furthermore, integrating the system with existing site management platforms (e.g., Autodesk BIM 360, Trimble Connect) is non-trivial due to data format heterogeneity and lack of unified interface standards.Future research will address these practical barriers by exploring scalable deployment strategies, adaptive redundancy for sensor fault tolerance, and API development for platform interoperability. Large-scale validations in infrastructure applications such as tunnel lining, bridge segment placement, and facade repair will also be prioritized to evaluate long-term performance in high-stress environments.

References

1. Frank M, Ruvald R, Johansson C, et al. Towards autonomous construction equipment-supporting on-site collaboration between automatons and humans[J]. International Journal of Product Development, 2019, 23(4): 292-308.

2. Zhang W, Ruttico P. Enhanced interactive AR bricklaying: elevating human-robot collaboration in augmented reality assisted masonry design and construction[J]. Architectural Intelligence, 2025, 4(1): 1-23.

3. Wu Q, Xue M, Lu B. Navigating the Divide: Balancing AI's Optimal Solution with the Appropriate Solution in Chinese Construction Management [J]. Journal of Current Social Issues Studies, 2025, 2(5): 276-290.

4. Mitterberger D, Dörfler K. Rethinking Digital Construction: A Collaborative Future of Humans, Machines and Craft[J]. Architectural Design, 2024, 94(5): 108-117.

5. Alexi E V, Kenny J C, Atanasova L, et al. Cooperative augmented assembly (CAA): augmented reality for on-site cooperative robotic fabrication[J]. Construction Robotics, 2024, 8(2): 28.

6. Rani S, Jining D, Shoukat K, et al. A human-machine interaction mechanism: additive manufacturing for Industry 5.0-design and management[J]. Sustainability, 2024, 16(10): 4158.

7. Nahmad Vazquez A, Garivani S, Dackiw J N. Decentralized, data-informed, robotic-based digital timber micro-factories[J]. Construction Robotics, 2024, 8(2): 24.

8. Vogel-Heuser B, Hartl F, Wittemer M, et al. Long living human-machine systems in construction and production enabled by digital twins: Exploring applications, challenges, and pathways to sustainability[J]. at-Automatisierungstechnik, 2024, 72(9): 789-814.

9. Zhang W, Ruttico P. Enhanced interactive AR bricklaying: elevating human-robot collaboration in augmented reality assisted masonry design and construction[J]. Architectural Intelligence, 2025, 4(1): 1-23.

10. Kunic A, Talaei A, Naboni R. Cyber-physical infrastructure for material and construction data tracking in reconfigurable timber light- frame structures[J]. Construction Robotics, 2025, 9(1): 11.