RESEARCH ARTICLE

Human–machine collaboration using gesture recognition in mixed reality and robotic fabrication

Alexander Htet Kyaw[1], Lawson Spencer[2] and Leslie Lok[1*]

Abstract

This research presents an innovative approach that integrated gesture recognition into a Mixed Reality (MR) interface for human–machine collaboration in the quality control, fabrication, and assembly of the *Unlog Tower*. MR platforms enable users to interact with three-dimensional holographic instructions during the assembly and fabrication of highly custom and parametric architectural constructions without the necessity of two-dimensional drawings. Previous MR fabrication projects have primarily relied on digital menus and custom buttons within the interface for user interaction between virtual and physical environments. Despite this approach being widely adopted, it is limited in its ability to allow for direct human interaction with physical objects to modify fabrication instructions within the virtual environment. The research integrates user interactions with physical objects through real-time gesture recognition as input to modify, update, or generate new digital information. This integration facilitates reciprocal stimuli between the physical and virtual environments, wherein the digital environment is generative of the user's tactile interaction with physical objects. Thereby providing user with direct, seamless feedback during the fabrication process. Through this method, the research has developed and presents three distinct Gesture-Based Mixed Reality (GBMR) workflows: object localization, object identification, and object calibration. These workflows utilize gesture recognition to enhance the interaction between virtual and physical environments, allowing for precise localization of objects, intuitive identification processes, and accurate calibrations. The results of these methods are demonstrated through a comprehensive case study: the construction of the *Unlog Tower*, a 36' tall robotically fabricated timber structure.

Keywords Human–Computer Interaction, Mixed reality, Gestural Tracking, Feedback Based Fabrication, Robotic Fabrication

1 Introduction

Mixed Reality (MR) serves as a bridge between the tangible physical environments and immersive virtual environments. Within the field of architecture, fabrication, and construction, this convergence holds significant potential for human–machine collaboration. Using MR, architects and designers can overlay digital blueprints directly onto physical geometries, enabling real-time instruction visualization (Rezvani et al., 2023). This paper aims to explore the collaboration between humans and machines to presents novel opportunities for fabrication efficiency, accuracy, and experience. The symbiosis of human expertise and machine feedback through MR processes presents a future that leads to new integrated workflows between human input, robotic fabrication and machine feedback within an immersive and phygital realm.

*Correspondence:
Leslie Lok
wll36@cornell.edu
[1] Rural-Urban Building Innovation Lab (RUBI), College of Architecture, Cornell University, Art, and Planning, Ithaca, NY 14853, USA
[2] Robotic Construction Laboratory (RCL), College of Architecture, Cornell University, Art, and Planning, Ithaca, NY 14853, USA

The term Mixed Reality encompasses both Augmented Reality (AR) and Virtual Reality (VR) within the Reality-Virtual (RV) Continuum. This continuum serves as a connection between real-world experiences and immersive virtual environments. (Milgram & Kishino, 1994). With the advancement of immersive technology and 3D user interfaces (3DUIs) in industry and academic research, the understanding of MR, as defined by Milgram and Kashino, has continuously evolved (Skarbez et al., 2021). In recent research, MR is often described as an environment-aware overlay of digital content on the physical world, enabling users to interact seamlessly with both environments (Speicher et al., 2019). To facilitate this interaction, MR systems employ an array of techniques, including spatial mapping, hand-tracking, eye-tracking, and auditory recording, collecting vital environmental and human physiological data. This amalgamation of the digital and physical in MR environments is further supported by advanced MR-enabled devices like the Microsoft HoloLens 2 and Meta Quest Pro, equipped with sensors, microphones, and cameras, enabling real-time monitoring of user behavior and changes in the physical environment (Microsoft, 2022).

Previous research using AR and MR workflows in the area of architectural fabrication have increased exponentially (Song et al., 2021b). Projects such as *Woven Steel*, *Timber De-Standardized,*, *Code-Bothy*, and many more have explored human interaction with digital instructions in MR through digital interfaces such as buttons, menus and/or fiducial markers such as QR codes and AruCo markers (Jahn et al., 2018a; Lee, 2022; Lok & Bae, 2022). These MR fabrication projects have focused on using human interactions with digital interfaces as the primary means to update the 3DUIs with new information. However, there exists an opportunity to directly incorporate human interaction with physical objects to update the 3DUI without needing digital interfaces.

The research integrates tactile interactions with physical objects through real-time gesture recognition as input to modify and update information in the digital environment. Through gesture recognition, the user's touch of a physical object could modify, update, or generate new digital information creating seamless stimuli between the physical and the virtual environments. By recording user gestures as they interact with physical objects, the three-dimensional user interface can automatically provide new information in real time. As a result, the virtual environment could respond dynamically to determine the real-time location of physical objects in the digital environment. This human machine collaboration can generate information such as localizing robotic tool paths, recognizing components, or measuring inaccuracies between the physical and the digital model. The real time generative data in the MR 3DUI allows the user to quickly respond to previous actions. The real time, feedback-based MR environment represents a cybernetic system whereby the outcome of interacting with a physical object(s) is taken as input for further action, thus creating a feedback loop until a desired condition is achieved.

The relationship between MR, gestural movement, digital twin, cybernetics, and human–computer interaction are used to help define systems of interaction between user and machine. From these relationships, the research presents three distinct *Gesture-Based Mixed Reality* (GBMR) fabrication workflows; a) *object localization*—registers the location of a physical object in the digital space, b) *object identification*—differentiates physical components using their digital parameters, c) *object calibration*—measures discrepancies between the physical object and associated digital geometry. Each of these three methods were used in six different tasks to construct the *Unlog Tower* (Fig. 1). The workflows derivative of this research presents new opportunities for human–machine co-creation within physical and virtual environments through MR in architecture and fabrication industries. The integration of tactile interactions plays a crucial role in allowing users to engage with digital data in a hands-on manner, effectively blending the physical and the virtual environments.

2 State of the art

Previous research projects have explored AR for Robotic fabrication to facilitate human–robot collaboration. "Implementation of an Augmented Reality AR Workflow for Human Robot Collaboration in Timber Prefabrication" proposes a user-friendly AR interface to visualize and manipulate robotic joint orientations, allowing users to send commands through a menu interface (Kyjanek et al., 2019). *Pop Up Factory*, employs an AR interface that allows users to manipulate digital control points of a wall assembly, thereby effecting the design of the 3D model used for subsequent robotic fabrication (Betti et al., 2019). Lastly, *[AR]OBOT,* employs an AR interface to visualize robotic operations in bricklaying applications. Users can plan the robotic movements by tapping on digital models of individual bricks (Song et al., 2021a). These projects have demonstrated the use of AR and MR interfaces for effective communication in robotic fabrication. However, these projects have primarily used AR and MR interfaces for the robotic fabrication of standardized work materials such as: foam blocks, bricks, or dimension lumber. The projects use AR and MR to engage with digital control points or menu interfaces. This paper demonstrates the potential to leverage gestural inputs for direct interaction with physical objects, providing spatial

Fig. 1 The *Unlog Tower*, Photo by Cynthia Kuo

data as parameters to enhance collaboration with robotic fabrication of both standard and non-standard materials.

Innovative fabrication research projects such as *Holographic Construction, Code-Bothy, Woven Steel, Bent*, and *Timber De-Standardized 2.0,* use interactive "buttons" for users to toggle between different sets of digital geometry which is visible in the 3DUI (Jahn et al., 2018a, 2019, 2020a; Lee, 2022; Lok & Bae, 2022). Though each of these projects use a *Microsoft Hololens* with *Fologram's* plug-in for *Rhino3d* and *Grasshopper,* the "buttons" can equally be interacted with one's mobile device. In each of these precedents, the "button" is a custom, pre-defined clickable digital object (either mesh or poly-surface). Thereby any change in the virtual interface is dependent on the user interacting with the select, pre-defined "buttons" or otherwise manipulating other digital geometry. *Holographic Construction* and *Code-Bothy* use digital "buttons" to toggle up and down between rows of bricks as they are laid (Jahn et al., 2020a; Lee, 2022). *Code-Bothy* has the added effect of color coordinating the amount of rotation per brick (Lee, 2022). *Woven Steel* and *Bent* exhibited several buttons to aid in the complex bending of tube steel and sheet metal (Jahn et al., 2018a, 2019). *Timber De-Standardized 2.0* developed menu list to visualize different aspects of an inventory of scanned irregular log meshes as well as cataloging and designing with the members through operations of slicing, indexing, baking, and isolating (Lok & Bae, 2022). Though these precedents offer an interaction between the user and the digital geometry, the interactions are limited to digital menus and buttons.

Other research projects such as *Timber De-Standardized 1.0, Augmented Feedback*, and *Augmented Vision* use various methods of AruCo markers for tracking,

physics simulation, and real-time scanning to create an active responsive environment between digital and physical objects (Lok et al., 2021; Goepel & Crolla, 2022; Jahn et al., 2022). In *Augmented Feedback,* AruCo makers were placed at nodal intersections of a bending-active bamboo grid-shell structure (Goepel & Crolla, 2022). AruCo marker tracking allowed users to digitize the locations of the markers and provide graphic feedback for all active users through the head mounted display (HMD). *Timber De-Standardized 1.0* utilized a physics simulation for fabricators to visualize and virtually "drop" irregular scanned meshes of logs till they found their resting point, which allowed for a precise alignment with its associated physical log (Lok et al., 2021). Finally, *Augmented Vision* uses the *Hololens 2* to track and scan the user's environment then display such information to inform the progress of constructing a minimal surface with strips of paper and/or bark (Jahn et al., 2022). These projects have demonstrated the capabilities of feedback-based MR using additional systems such as AruCo markers, scanned meshes, and simulation.

Additionally, the accuracy of AR/MR platforms presents a significant challenge in many of these AR/MR fabrication workflows. The accuracy of the fabrication instructions provided to users depends on the precision of the system. As a result, several studies have been conducted to assess the accuracy of AR/MR systems. Researchers have investigated the use of AR for assembling metal pipes (Jahn et al., 2018b), weaving bamboo structures (Goepel & Crolla, 2020), and constructing complex wall systems with bricks within a tolerance of ± 20 mm (Jahn et al., 2020b). Moreover, there have been research efforts aimed at improving the accuracy of AR/MR systems. The paper, "Augmented Reality for High

Kyaw *et al. Architectural Intelligence*   (2024) 3:11

Page 4 of 18

Precision fabrication of Glued Laminated Timber Beams", has explored the use of multiple QR codes to achieve a tolerance below 2 mm with the Microsoft *HoloLens* 2 (Kyaw et al., 2023). The results of this study indicate that AR/MR systems have the potential to be used for high precision applications, such as assisting in robotic fabrication and accurate quality control.

3 Aim and Objectives

The research presented in this paper investigates various applications where several GBMR workflows can leverage tactile feedback to enrich the user experience when interacting with both physical and virtual items. The paper demonstrates how tactile interactions can be used to visually enhance the user's perception with additional digital information when manipulating physical objects. The research exhibits how the three described GBMR workflows can create a more immersive and fluid interaction methodology that capitalizes on the human's natural sense of touch, enabling users to physically feel and interact with the virtual environment in a tangible way. While previous MR projects have focused on using menus, AruCo markers, scanned meshes, and simulations to interact with digital geometries, this project investigates the potential of incorporating user's tactile interaction with physical objects as an input to update the 3DUI.

This research has developed 6 experiments to test 3 GBMR fabrication workflows to enhance tactile interactions by generating geometry relative to physical objects, localizing robotic tool paths, recognizing discrete components according to parameters such as height and length, and measuring inaccuracies between the physical and the digital models. The paper will first present the tools and software of the method, which will then be followed by the three GBMR workflows used to fabricate

the *UnLog Tower*: a) *object localization,* b) *object identification,* and c) *object calibration. Object localization* was used to determine the log geometry work object and the toolpath placement for robotic fabrication (Method 4.1) (Fig. 2). *Object identification* is utilized to identify physical components and display intuitive step-by-step assembly instructions (Method 4.2). *Object calibration* is employed to ensure the adjustment of jigs and the connection of panels match the digital model (Method 4.3).

Each of these workflows will demonstrate new methods in MR research whereby physical stimuli can become a generative tool to interact and inform MR fabrication in real-time. Through gestural interaction, our research endeavors to redefine the boundaries between the physical and virtual environments. By showcasing their application in the construction of the *Unlog Tower,* these workflows demonstrate potential to optimize fabrication processes, enhance assembly efficiency and instruction, thereby contributing to an advancement within the field building construction.

4 Methods

Through computer vison and gestural recognition algorithm, the following studies were conducted with a Microsoft *HoloLens 2* and *Fologram*, a AR/MR plug-in for *Rhino3D* and *Grasshopper* (Fologram Pty Ltd, 2021; Robert McNeel & Associates, 2022; Rutten, 2022). The near depth sensing camera on the Microsoft *HoloLens 2* is used for articulated hand tracking (AHAT). AHAT is a computer vision algorithm that tracks the movement and gestures of the user's hand, independent from the visible light cameras (VLCs) used for simultaneous locating and mapping (SLAM). The articulated hand tracking system recognizes and records twenty-five 3D joint positions and rotations, including the wrist, metacarpals, proximal,
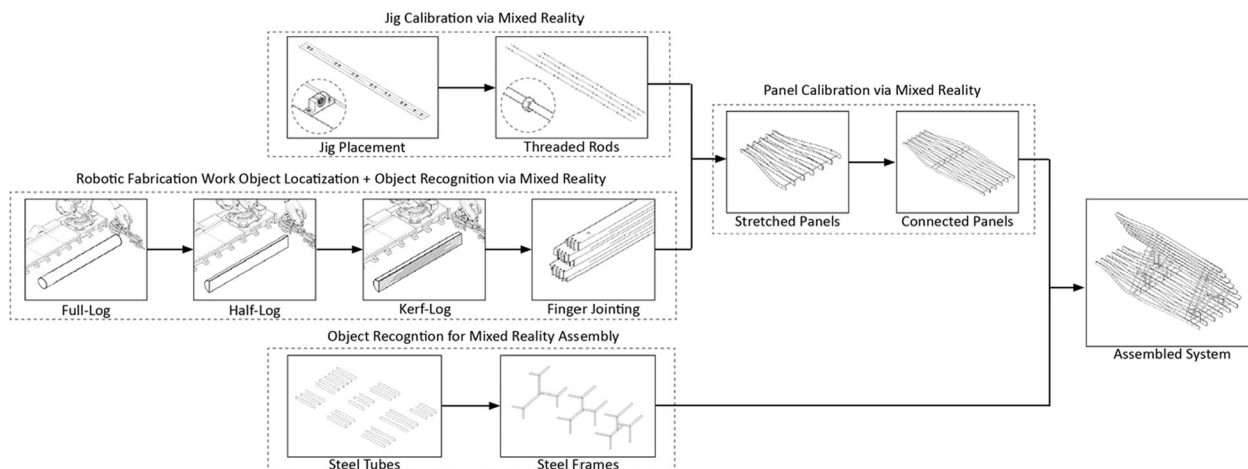

Fig. 2 Workflow diagram outlining the various assembly and fabrication process

distal, and fingertip joints (Ungureanu et al., 2020). This data is live streamed from the *HoloLens 2* device to *Rhino3D* and *Grasshopper* via *Wi-Fi*. The Microsoft AHAT API provides access to the built-in gestural recognition algorithm of the *HoloLens 2*, enabling the utilization of its advanced capabilities for hand tracking purposes. The joint configuration and orientation obtained from AHAT can facilitate the estimation of hand poses, such as pinching, tapping, or poking (Taylor et al., 2016).

This study focuses on the use of pinching as the primary mode of gestural interaction by the user. The pinching gesture is recognized when the thumb tip and index fingertip are in close proximity (Fig. 3). Therefore, a device capable of AHAT programming is imperative for gesture recognition and therefore is integral to the GBMR workflows. Gestural recognition plays an important role in tracking tactile interactions and serves as the input for human–machine collaboration in GBMR workflows.

4.1 Object localization
The *Unlog Tower* exhibits robotically kerfed timber round woods that have been stretched along two threaded rods to form panels through a similar method exhibited at the *Unlog* pavilion at University of Virginia (Lok et al., 2023). Logs are irregular geometries that are comprised of knots and are sometimes curved, but can nonetheless be abstracted to a cylinder in most cases. Six ash logs with minor deformations were used to construct the
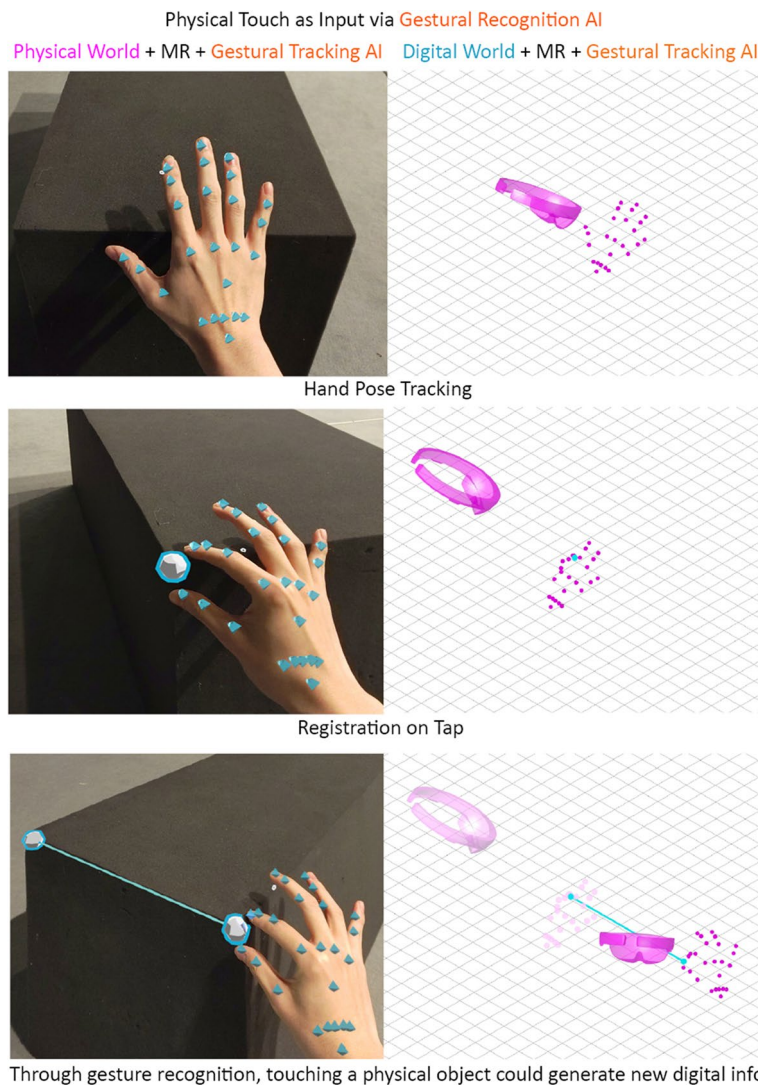

Fig. 3 Digital twin of *HoloLens 2* headset location, joint configuration, and orientation from AHAT (Articulated Hand Tracking); visualized through headset (left); visualized through *Rhino3D* and *Grasshopper* (right)

tower, each log was first cut in half and then robotically kerfed. To localize the robot targets and cut the log in half using a 6-axis robotic arm with a 5hp bandsaw end-effector, *object localization* method was employed. The user placed three points at both ends of the log to create two individual circles that generated a cylindrical mesh which was superimposed with the physical log (Fig. 4). Each point was created by the user pinching their right-hand index finger to their thumb. This feedback mechanism provides the user with a visual confirmation of the digitization process by displaying a point for each gesture recorded. From the cylindrical mesh, a surface was generated in the middle of the cylinder whereby the robot tool path could be derived from the robot targets at either end of the surface using *Robot Components* (Deetman et al., 2023), a robot programming plug-in for ABB robots in *Grasshopper* that is then copied into *Robot Studio,* an ABB software for programming ABB robots (ABB, 2023).

Once the log was cut in half, one half of the log was rotated 90° and remounted in the robot cell. According to the structural requirements for the *Unlog Tower,* the cross section of each board was to be no less than $5''$ by $0.75''$. Figure 5 demonstrates the process whereby the user would locate the half log in the robot cell by placing three points; two at one side of the half log to determine the diameter and one at the opposite end to determine the length of the half log (Fig. 5). After the log geometry was defined, the user set the location of the cut geometry by placing a point on the profile of the log (Fig. 6).

The MR workflow offered the user ongoing feedback throughout the process by performing a validation to determine whether the cut geometry falls within the boundary of the log. In the event that the cut geometry was placed outside the log or was situated too close to the log mount, a red notation with a cross mark was displayed within the 3DU (Fig. 7a and b). The user responded to the alert and adjusted the location of the cut geometry until a satisfactory outcome was achieved, represented by a green notation (Fig. 7c). The fabricator was to check the location of the cut surfaces within the log to ensure that the boards met the minimum cross-sectional requirements without any of the cut surfaces colliding with the $4'' \times 4''$ log mounts. The *object localization* workflow allows users to define points in the digital space that represent the physical log for work-object localization during robotic fabrication (Fig. 8).
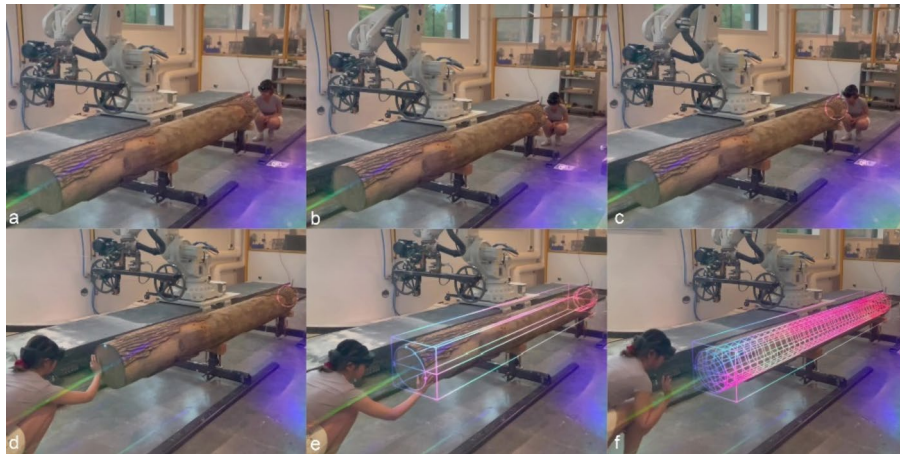

Fig. 4 *Object localization* is used to generate the location of a cylinder according to the diameter(s) of the log to automate the placement of the robotic toolpaths

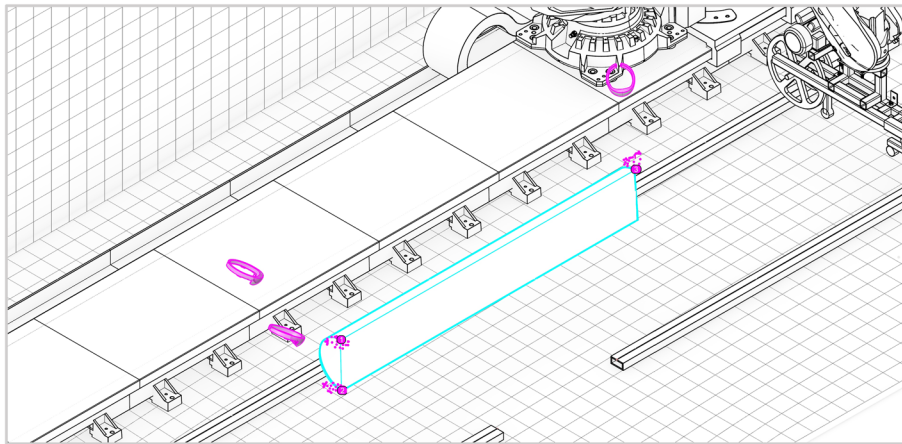**Fig. 5** *Object localization* is used to determine the work object placement for robotic fabrication

Fig. 6 Gestural inputs are used to register the location of a physical object in the digital space for robotics
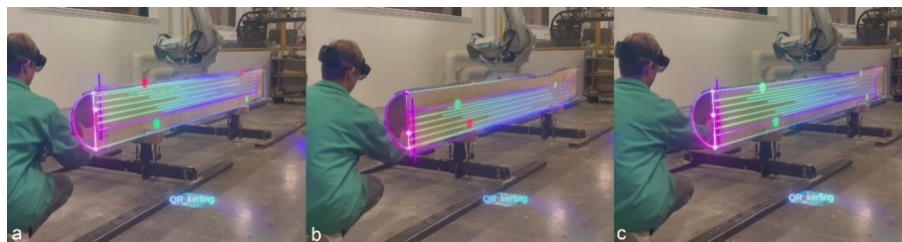

Fig. 7 *Object localization* is used to determine the placement of the toolpath for robotic fabrication
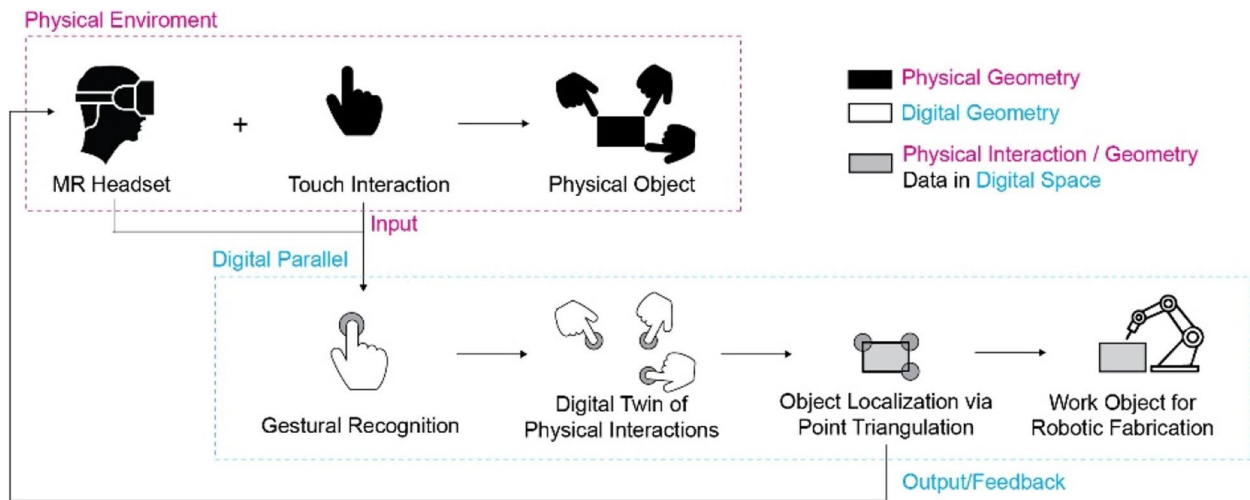

Fig. 8 *Object localization* system diagram describing how user interactions physical objects are used to create digital data through gestural recognition

An ABB IRB 6700 on a 4200 mm external linear track was used to cut each half log into robotically kerfed, bending-active panels (Fig. 9).

4.2 Object identification

Object identification was used to differentiate between self-similar physical components and display intuitive step-by-step assembly instructions. After the half logs

Fig. 9 6-axis robotic arm with a 5hp bandsaw end-effector cutting a log after object localization

have been robotically kerfed, they are set aside and prepared for finger jointing. The finger joint template not only includes an outline for the finger joints, but also an outline for the hole that the threaded rod would pass through. Because of the parametric design of the kerfed timber panels for the *Unlog Tower,* the finger joint locations are staggard between adjacent boards within each half log (Fig. 10).

In order to correctly mark the location of the finger joints and the location for the threaded rod holes in each board layer, GBMR was employed for *object identification.* Each board layer had a defined thickness of 0.75 inches. Therefore, the height of the virtual templates were at intervals of 0.75 inches (e.g., Layer 1: 0.75 inches, Layer 2: 1.5 inches, Layer 3: 2.25 inches, and so on). *Object identification* was specifically used to identify the board layer that the user was working on to display the corresponding virtual template location. The workflow determines the corresponding virtual template to display by comparing the height of the user-defined point with height of the virtual templates from the ground plane

(Fig. 11). For instance, if the user specifies a gestural point positioned 1.43 inches above the ground, the system will match this value with the nearest layer height within a virtual template. In this scenario, the system will present layer 2 as the closest match, positioned at 1.5 inches above the ground, which closely corresponds to the input of 1.43 inches. The virtual template had an added notation that visually communicated to the user which layer they were working on, so that the user could be sure that the physical template was appropriately placed. The finger joints were cut with an oscillating saw and drill, while the holes for the threaded rods were drilled with a hole saw (Fig. 12). This *object identification* workflow allows for fluid transition between the physical world and the digital overlays, where users can simultaneously navigate digital instruction and fabricate physical geometries.

Additionally, *object identification* was used to index and coordinate between self-similar parts. Through gestural recognition, tactile interactions with physical geometries were recorded as digital points. These points were sorted in the order of registration to calculate the distance
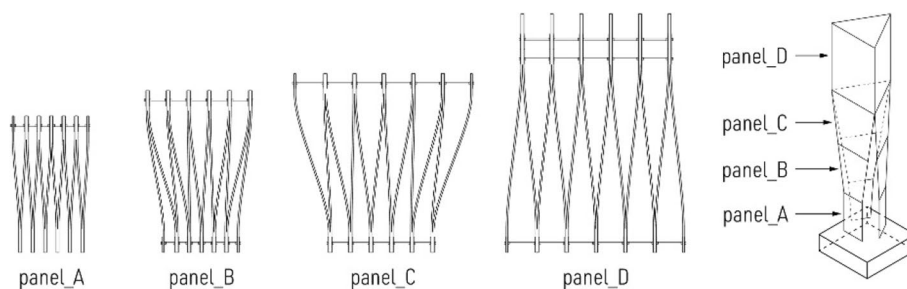

Fig. 10 Staggered board layers depending on kerf panel geometry and parameter

Fig. 11 *Object identification* is utilized to identify physical components and display intuitive step-by-step assembly instructions

Fig. 12 Robotically Kerfed logs with finger joints and threaded rod holes

between each gesture. This distance parameter was used to match the corresponding digital instruction for the user. This human–machine collaboration was exhibited in the fabrication of the reciprocal tube steel frames in the *Unlog Tower*. To brace the kerfed wood panels, the interior of the tower exhibited 3 sets of steel tube frames.

Due to the custom design of the steel tube frames, there were nine unique tube lengths amongst 54 total steel tubes (Fig. 13). Seven of the nine steel tube lengths were: 17.27 inches, 18.82 inches, 22.28 inches, 23.20 inches, 24.83 inches, 27.72 inches, and 32.93 inches. After the steel tubes were cut to length, *object identification* was employed to index the tube steel according to their length and communicate the location of the tube steel in the digital model(s) (Fig. 14a and c). By placing a point at either end of the of the tube steel through gesture recognition, the user would define the length of the object, which was checked against a list of tube steel lengths predetermined in the digital model. If the value between the user defined length and a predefined length was within tolerance (see Table 2 in the Results), the 3DUI displayed the corresponding digital information to the user through notation and two coordination models that visually indicated the location of the tube steel within the overall structure and highlights the selected member from to blue to red. The coordination model on the left (Fig. 14 b and d) illustrated at 1:1 scale the tube steel location within the associated tube steel frame and the coordination model on the right (Figs. 14a, b, 15c, and 14d) illustrated at 1:10

Fig. 13 Reciprocally framed tube steel in the *UnLog Tower,* photo by Cynthia Kuo.
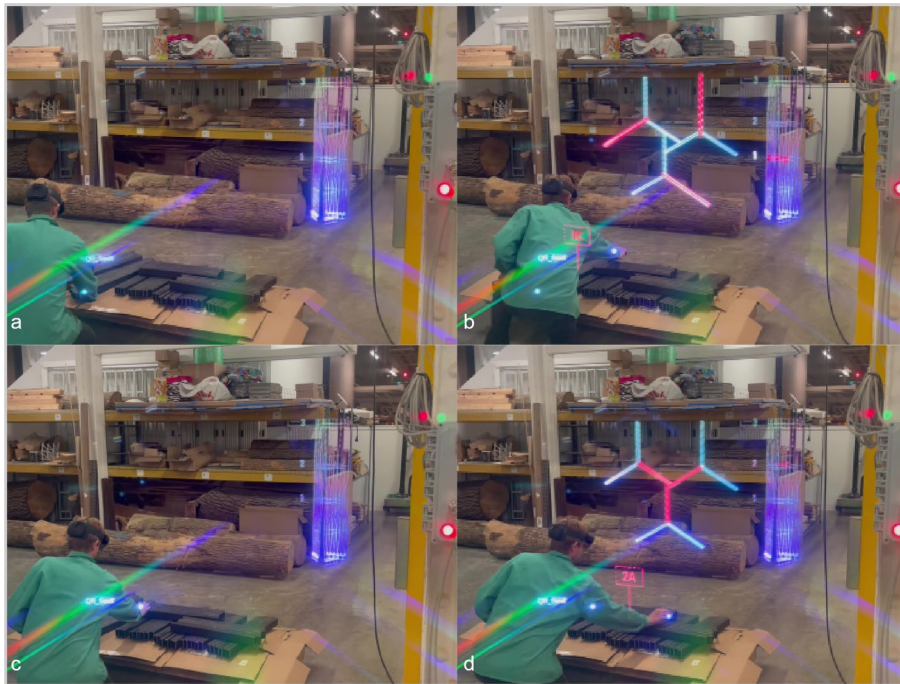

Fig. 14 *Object identification* is utilized to identify physical components and display part to whole assembly instructions

scale a virtual model of the *Unlog Tower* with the location of the tube steel within the whole model. By using predetermined distances and gestural recognition, *Object Identification* was used to pair digital assembly instructions with the identified physical object (Fig. 15).

4.3 Object calibration
In order for the kerfed logs to splay out into panels, the threaded rods had to have pre-located hex nuts appropriately placed to ensure that each board member would be in the correct location. In the GBMR workflow, *object calibration* was employed to place the hex nut locator correctly along a plywood jig. The hex nut locator was 3D printed with PLA to firmly hold each hex nut when it was screwed into the plywood board. A digital twin was created for each hex nut locator. This 3D printed hex nut locator had a handle that protruded 0.25 inches with a thickness of 0.125 inches. When the user pinched the handle on the hex nut locator, *object calibration* would use gesture recognition to continuously track this movement, thereby synchronizing the digital geometry with the physical. As the physical object moved closer to the

goal position, the notation would transform from red to yellow to green once the physical was properly located (Fig. 16).

This workflow represented a cybernetic system in which the adjustment of the physical locator position would generate new virtual feedback for the user, thus creating a feedback loop until the desired condition was attained. The desired condition was achieved when the digitized physical location of the hex nut locator was within a tolerance of 0.125 inches. This was indicated

to the user via the notation system where the red or yellow cross turned into a green tick. The MR system would instruct the user to move onto the next hex nut locator only after the previous hex nut locator was correctly placed via gesture recognition. After all the hex nut locators were properly placed, a threaded rod was screwed through the jig (Fig. 17).

For the panel assembly, the robotically kerfed logs were splayed out along two threaded rods with pre-located hex nuts as was done in the *Unlog* pavilion (Lok et al.,
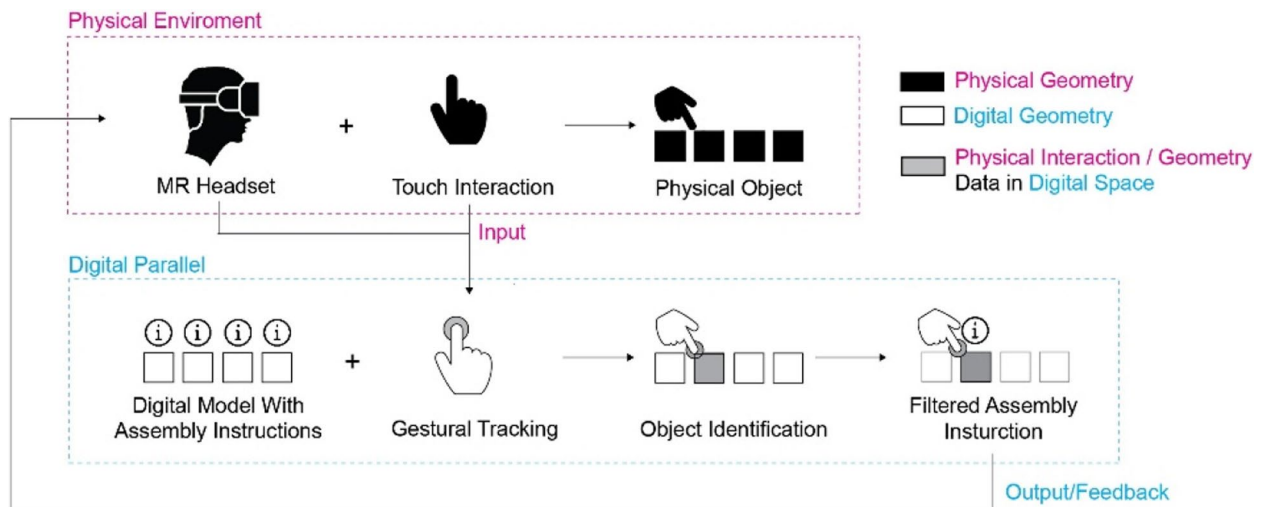

Fig. 15 *Object identification* system diagram describing how digital assembly is filtered through object identification via gestural recognition

Fig. 16 *Object calibration* is employed to ensure the hex nut locators are adjusted to match the digital model. As the physical hex nut locator moves closer to its digital position, the notation would transform from red to yellow to green

Fig. 17 After all the hex nut locators were properly placed, a threaded rod is screwed through jig
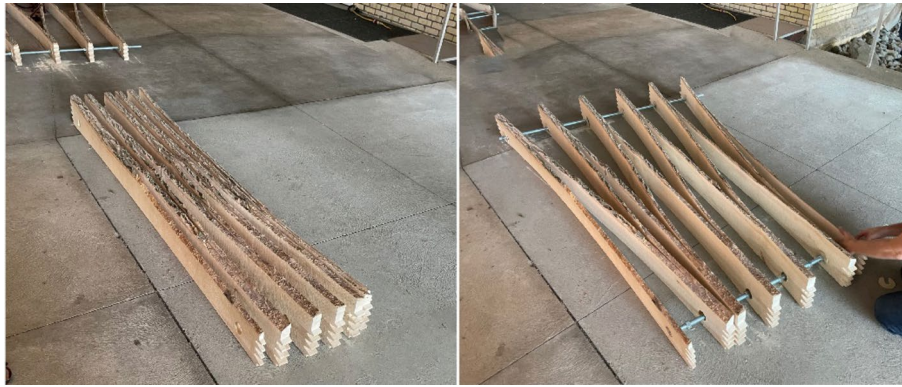
Fig. 18 Transformable material system at two phases: Collapsed kerf log (left) and Stretched kerf log (Right)

2023) (Fig. 18). Temporary custom slip washers were placed between the hex nut and the board to ensure that the boards would keep their position until joined into larger prefab components with steel slip washers. Once the panels were joined together in larger prefab components, *object calibration* was used to check the location of each board as they were fixed into location (Fig. 19). This quality control step aligned a digital model of the goal geometry to the physical panel using the placement of a QR code. The physical location of the boards were determined by using GBMR to place a point at the center of the finger joint location on each board, which was automatically checked against the closest digital board from the 3D model. The deviation between the GBMR input board location and the digital board allowed for a 0.125″ tolerance. A red cross notation indicated that the deviation was outside the tolerance, otherwise a green check notation would appear indicate that the board was correctly placed.

Object Calibration, as a quality control step, ensured that the parametrically defined wall panels were properly calibrated into larger prefab wall elements that were then transported to the site for assembly (Fig. 20). The utilization of gestural recognition allowed the machine to record user's tactile interaction with physical objects. By measuring the distance between the physical and the digital objects, the machine can understand the fabrication tolerances in real-time and provide an immediate visual feedback to the user (Fig. 21).

5 Results and Discussion

The implementation of gesture recognition for GBMR was incredibly useful for the fabrication of irregular and parametrically defined building components exhibited in the construction of the *Unlog Tower*. The prefab wall panels were attached to the tube steel reciprocal frames on site and lifted onto the foundation with a boom forklift (Fig. 22). The *Unlog Tower* was on display for 6 months until it was deinstalled in March of 2023.

Gestural recognition in MR fabrication workflows allowed users to define physical objects without the arduous placement of AruCo markers. The *object localization workflow* demonstrates that gesture recognition can be employed to locate robot work object data (Fig. 8). However, the utilization of gesture recognition assumes a certain level of dexterity on the part of the user, as the data is dependent on the fidelity and accuracy of the user's fingers. During the experiment, no issues were encountered regarding the fidelity of the user's finger. Since robotic fabrication was utilized for kerfing logs, the workflow achieves its intended outcome as long as the work object remains within the
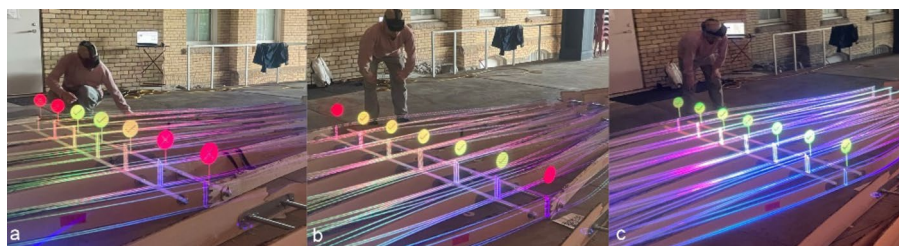

Fig. 19 *Object calibration* is employed for quality control of prefab wall components
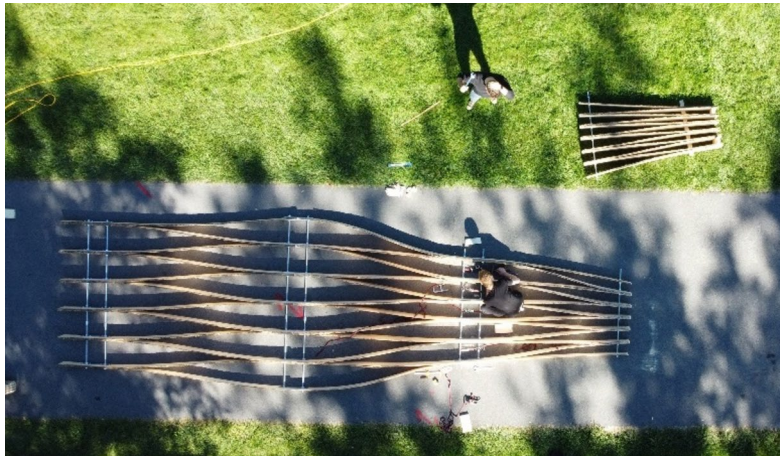
Fig. 20 Aerial of the kerfed panels assembled into larger wall components, photos by Cynthia Kuo.
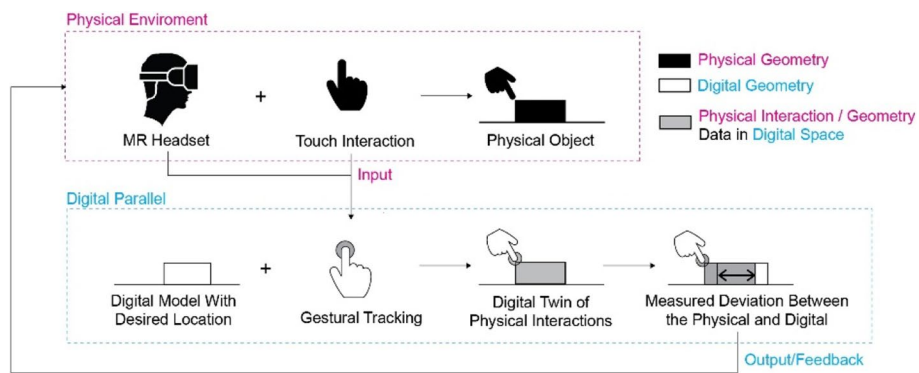

Fig. 21 *Object calibration* is employed to ensure the adjustment of jigs and the connection of panels match the digital model

Fig. 22 Aerial photograph of the *Unlog Tower* lifted on the foundation pad with a boom forklift, photo by Cynthia Kuo

Kyaw *et al. Architectural Intelligence*   (2024) 3:11

Page 14 of 18

width of the robotic bandsaw. However, robotic fabrication processes such as milling might require higher accuracy. Future studies will investigate how the *object localization* workflow can be modified for robotic fabrication procedures that require higher tolerances. Alternatively, improvements in the AHAT, articulated hand tracking, on the *Microsoft HoloLens 2* would also increase the accuracy of the overall system and the resolution of the work object placement.

The research also describes the potential of using gestural tracking for *object identification* whereby the user's hands can be intuitively used to index and coordinate assembly of self-similar parts based upon predefined parameters (Fig. 15). The allowable range of a user positioned points through gesture recognition is defined as the *gestural input tolerance.* As object parameters are relative to one another, the gesture input tolerance is also relative to adjacent parameters within a list, so the lower limit of the *gesture input tolerance* for a specific object x_n can be found by calculating the midpoint between the predefined parameters of the preceding object x_{n-1} and object x_n. The upper limit of this range can be determined by calculating the midpoint between the predefined parameters of subsequent object x_{n+1} and object x_n (Eq. 1).

corresponds with layer 3. The lower limit of the gesture input tolerance for layer 2 is calculated by finding the midpoint between the heights of layer 1 and 2, while the upper limit is the midpoint between layer 2 and 3.

Another value that was used to measure the robustness of the system is the *identification threshold*. The *identification threshold* represents the smallest allowable deviation the user's gestural input can have before the system identifies the wrong object. The *identification threshold* of object x_n can be calculated by finding the lesser difference between the geometry parameter of object x_n and that of its preceding object x_{n-1} and subsequent object x_{n+1} (Eq. 2). The *identification threshold* is negative if the preceding object (x_{n-1}) has a smaller difference. The *identification threshold* is positive if the subsequent object (x_{n+1}) has a smaller difference. If the two values are equal, then the *identification threshold* has both positive and negative value. In this experiment, the *identification threshold* for all board layers is ± 0.375 inches. This means any gestural input deviating by more or less than 0.375 inches from the object's layer height will result in a misidentification. During the board layer identification experiments, the system was able to accurately identify all corresponding layers without any errors for the *identification threshold.*

$$f(x_n) = \begin{cases} -min(|x_{n-1} - x_n|, |x_n - x_{n+1}|), if |x_{n-1} - x_n| < |x_n - x_{n+1}| \\ +min(|x_{n-1} - x_n|, |x_n - x_{n+1}|), if |x_{n-1} - x_n| > |x_n - x_{n+1}| \\ \pm min(|x_{n-1} - x_n|, |x_n - x_{n+1}|), if |x_{n-1} - x_n| = |x_n - x_{n+1}| \end{cases} \qquad (2)$$

$$\left[\frac{x_{n-1} + x_n}{2}, \frac{x_n + x_{n+1}}{2} \right] \qquad (1)$$

In the first *Object Identification* experiment, gestural input was used for board layer identification. In this context, the *gesture input tolerance* refers to the acceptable range within which a user's gestural inputs must fall for the system to accurately identify the corresponding board layer. (Table 1). For example, the *gesture input tolerance* for layer 2 is between 1.175 and 1.825 inches. This means any gestural input falling below the lower limit of 1.175 inches will correspond to the virtual template of layer 1, while any input above the upper limit

The second experiment in *Object Identification* recognizes distinct tube steel types by utilizing varying lengths of the members as geometry parameters. In contrast to the initial experiment, which focused on incremental differences in layer height, this experiment involves tube steel length variations with non-uniform differences among individual members. Due to this non-uniform varying, the *gesture input tolerance* between each member was drastically different. For example, Type D has a *gestural input tolerance* between 22.74 inches to 24.015 inches which is a range of 1.275 inches, and Type G has a *gestural input tolerance* between 30.325 to 35.535 inches which is a range of 5.21 inches (Table 2). As a result, it is more likely for a user's

Table 1 Gestural Input Tolerance and Identification Threshold for Uniform Board Layer Identification

Board Layer No	Layer Height	Gestural Input Tolerance	Identification Threshold
Layer 1	0.75 inches	0 to 1.175 inches	± 0.375 inches
Layer 2	1.5 inches	1.175 to 1.825 inches	± 0.375 inches
Layer 3	2.25 inches	1.825 to 2.25 inches	± 0.375 inches
Layer 4	3.0 inches	3.0 to 3.37 inches	± 0.375 inches
Layer 5	3.75 inches	3.37 to 4.5 inches	± 0.375 inches

Kyaw *et al. Architectural Intelligence*   (2024) 3:11

Page 15 of 18

Table 2 Gestural Input Tolerance and Identification Threshold for Steel Tube Identification

Steel Tube Type	Steel Tube Length	Gestural Input Tolerance	Identification Threshold
Type A	17.27 inches	16.495 to 18.045 inches	+ 0.775 inches
Type B	18.82 inches	18.045 to 20.55 inches	- 0.775 inches
Type C	22.28 inches	20.55 to 22.74 inches	+ 0.46 inches
Type D	23.20 inches	22.74 to 24.015 inches	- 0.815 inches
Type E	24.83 inches	24.015 to 26.275 inches	- 0.815 inches
Type F	27.72 inches	26.275 to 30.325 inches	- 1.445 inches
Type G	32.93 inches	30.325 to 35.535 inches	- 2.605 inches

gestural input to fall out of bound for Type D compared to Type G. However, Type D can be misidentified as either Type C or Type E. The *identification threshold* can be calculated if it is more likely for the system to identify Type C as Type C or Type E. The *identification threshold* of Type D is -0.46. In reference to Eq. 2, the negative value of the *identification threshold* was attributed to smaller differences with the preceding object. Therefore, the likelihood of the system misclassifying Type D as Type C was higher. Throughout the experiment, there were two instances of error recorded during the five documentation trials. Both of these errors occurred when the system mistook Type D as Type C.

When comparing the two *object identification* experiments, the *identification threshold* in the first experiment had consistent value of 0.375 inches. While this value is smaller than the *identification threshold* of Type D in second experiment, there was no error recorded in the first experiment. However, it is also important to note that in the first experiment, the user only needed to input one gestural point for the system to read the layer height. In the second experiment, the user needed to input two gestural points to register the tube steel length. Registering two points means that the identification through gestural recognition could have an increased possibility of error. Future research will conduct a precision study on how the number of gestural points can lead to a higher discrepancy. The results also indicate that type of geometry parameters has a significant role in the performance of an *object identification* workflow using the GBMR method. Currently, the *object identification* method utilizes the varying lengths and heights of components as the parameter. Future studies could incorporate other geometric parameters such as the boundary geometry or volume in the workflow.

The research underscores a critical aspect of visual feedback of human–machine collaboration by developing visualization strategies for various fabrication tasks. For true collaboration to exist, there must also be a mutual understanding between the user and the system. The machine must be able to comprehend the user's input, and the human must also be able to understand the machine's outputs. Utilizing gestural recognition, the machine is capable of capturing and processing interactions initiated by users. Subsequently, the machine generates outputs that enhance the user's tactile interaction by providing real-time visual feedback.

In the case of the *object localization* workflow, the accuracy of the gesture recognition is limited to the user's finger precision. The tactile interaction is enhanced with visual feedback by displaying a sphere at the location of the placement point to verify the physical input. Preliminary experiments have recorded users recalling their tactile interactions when they notice discrepancies displayed in the visual feedback. This visual feedback enhancement enables users to see errors between physical action and the digital output.

Integrating visual perception also plays a crucial role in the *object identification* workflows, where 3D drawings and instructions are dynamically updated based on the user's tactile interactions. During the kerf panel fabrication, we noticed that it was challenging to identify if a task is registered without clear labeling on each panel layer. Specific labels and colors have been added as a form of visual feedback to draw attention to updated information. During the steel frame fabrication, the change in color highlighting the selected member allows the user to confirm that their object identification was successful.

Finally, the *object calibration* workflow showcases a synchronized method for users to link physical objects with their digital twins (Fig. 21). The threaded rod test was unique in that the user could pinch the hex nut locator while moving the physical object. Visual feedback can was used to enhance tactile interaction through color coordination. For example, the instructions can shift colors from red, to yellow, to green in response to the user's physical inputs, effectively signaling to anticipate when they would be close to the goal location. Users have reported that the visual feedback provides them with more confidence in their actions during the fabrication process. Through the employment this workflow,

all 24 threaded rods of the *Unlog Tower* were successfully fabricated as intended. The second *object calibration experiment* with the panel quality control demonstrated that some objects are too heavy or cumbersome to pinch while moving. For that reason, the second test demonstrated the use of gesture recognition to iteratively define critical points until the physical geometry aligned with the digital model.

With the development of Gesture-Based Mixed Reality workflows for object localization, identification, and calibration, the research advances current fabrication processes by enabling real-time feedback through tactile interaction. By enabling direct interaction with three-dimensional holographic instructions, the need for two-dimensional drawings in other fabrication processes is eliminated, allowing for a more interactive and tactile engagement with the fabrication tasks. Without relying on physical measurement tools such as measuring tapes or rulers associated with common fabrication practices, the method can handle complex, parametric, and irregular geometries while accounting for fabrication errors.

This workflow can also have a drastic impact on the industry and the manpower involved in the fabrication process. By changing the nature of how fabrication drawings and technical documentation are produced, the workflow makes it easier for teams to understand and follow complex fabrication instructions. Previously, reading technical drawings would be limited to those with specialized training in architecture or construction. While using a mixed reality headset still requires training, it is still a lower barrier of entry into certain fabrication tasks. The use of interactive fabrication instructions and real-time feedback opens up opportunities for experts and nonexperts to fabricate highly customized and unique geometries. The research also presents opportunities for fabricators to develop future projects that employ this method to coordinate and educate subcontractors on the construction of parametric components with discretized or self-similar parts. The use of gesture recognition and MR in fabrication projects is not just about improving human–machine collaboration; it's also about enhancing human–human collaboration.

6 Conclusion

The future potential of using gesture recognition in MR fabrication projects is enormous. The presented research not only demonstrates that real time feedback through gesture recognition is imperative for advanced MR fabrication projects, but it can also be used in robotics, geometry creation, object indexing, model coordination, interactive digital twin, and complex quality control. In the age of automation, the research highlights the importance of integrating human

interaction into machine processes. The research presents a concurrent bi-directional human–machine collaboration workflow. The focus isn't solely on humans giving commands to machines or machines directing humans. Instead, it is about fostering a deeper understanding and synergy between both entities, working collaboratively to improve and optimize outcomes. The integration of tactile interaction and gesture recognition embodies this collaboration, enabling users to not only interface with the digital environment but also to effectively collaborate with machine generated information.

The insights gained from the experiments conducted in this study pave the way for future explorations, offering innovative approaches to integrate physical stimuli as generative tools for MR fabrication in real-time. Future investigations will seek to improve the accuracy of this method for high precision fabrication projects and explore the potential of incorporating a wider range of gestures, such as "tap", "poke", and "pinch". Additionally, the development of a user-controlled interface to manage recognized gestures, enabling actions such as enable/disable or undo, will further refine the collaborative dynamics between the user and the system.

This research demonstrates how gesture-based mixed reality workflows can provide a tangible interface to simultaneously interact with both physical objects and digital content within mixed reality environments. By leveraging tactile interactions, the workflow redefines the boundaries between the physical and digital domains, ultimately pushing the limitation of immersive technology for feedback-based human–machine collaboration in construction and related fields. The three GBMR workflows exhibited in this paper demonstrate the various applications for the real-time feedback-based fabrication and assembly of the *Unlog Tower*. This phygital experience offers a whole series of future applications investigations in the field of Mixed Reality fabrication and Human–Machine co-creation.

Authors' contributions
Conceptualization and Methodology: Alexander Htet Kyaw, Lawson Spencer, and Leslie Lok; Formal analysis and investigation: Alexander Htet Kyaw; Principal Investigator and Funding acquisition: Leslie Lok; Co-writing: Leslie Lok and Lawson Spencer.

References

ABB. (2023). *RobotStudio* [RAPID; Windows]. ABB.

Betti, G., Aziz, S., & Ron, G. (2019). Pop Up Factory: Collaborative Design in Mixed Rality Interactive live installation for the makeCity festival, 2018 Berlin. *Blucher Design Proceedings*, 115–124. https://doi.org/10.5151/proceedings-ecaadesigradi2019_425.

Deetman, A., Wannemacher, B., & Rumpf, G. (2023). *Robot Components* (1.5.1) [C++; Windows]. Robot Studio.

Fologram Pty Ltd. (2021). *Fologram* (Version 2020/02/15) [Windows]. Fologram Pty Ltd.

Goepel, G., & Crolla, K. (2020). Augmented Reality-based Collaboration—ARgan, a bamboo art installation case study. *Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia*, 313–322. https://doi.org/10.52842/conf.caadria.2020.2.313.

Goepel, G., & Crolla, K. (2022). Augmented Feedback: A case study in Mixed-Reality as a tool for assembly and real-time feedback in bamboo construction. In K. Dörfler, S. Parasho, J. Scott, B. Bogosian, B. Farahi, J. L. García del Castillo y López, J. A. Grant, & V. A. A. Noel (Eds.), *ACADIA 2021: Toward Critical Computation* (pp. 232–237). ACADIA. https://doi.org/10.52842/conf.acadia.2021.232

Jahn, G., Newnham, C., & Beanland, M. (2018a). Making in Mixed Reality. Holographic design, fabrication, assembly and analysis of woven steel structures. *Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture.*, 88–97. https://doi.org/10.52842/conf.acadia.2018.088.

Jahn, G., Newnham, C., & Beanland, M. (2018b). Making in Mixed Reality. Holographic design, fabrication, assembly and analysis of woven steel structures. *Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture.*, 88–97. https://doi.org/10.52842/conf.acadia.2018.088.

Jahn, G., Newnham, C., & Berg, N. (2022). Depth Camera Feedback for Guided Fabrication in Augmented Reality. In Dr. D. Aviv, H. Jamelle, R. Stuart-Smith, & Dr. M. Akbarzadeh (Eds.), *Proceedings of the 42nd Annual Conference of the Association for Computer Aided Design in Architecture* (ACADIA). ACADIA. https://papers.cumincad.org/cgi-bin/works/paper/acadia22_684

Jahn, G., Newnham, C., van den Berg, N., Iraheta, M., & Wells, J. (2020a). Holographic Construction. In C. Gengnagel, O. Baverel, J. Burry, M. Ramsgaard Thomsen, & S. Weinzierl (Eds.), *Impact: Design With All Senses* (pp. 314–324). Springer International Publishing. https://doi.org/10.1007/978-3-030-29829-6_25

Jahn, G., Newnham, C., van den Berg, N., Iraheta, M., & Wells, J. (2020b). Holographic Construction. In C. Gengnagel, O. Baverel, J. Burry, M. Ramsgaard Thomsen, & S. Weinzierl (Eds.), *Impact: Design With All Senses* (pp. 314–324). Springer International Publishing. https://doi.org/10.1007/978-3-030-29829-6_25

Jahn, G., Wit, A. J., & Pazzi, J. (2019). [Bent] Holographic handcraft in large-scale steam-bent timber structures. *Proceedings of the 39th Annual Conference of the Association for Computer Aided Design in Architecture* (ACADIA), 438–447. https://papers.cumincad.org/data/works/att/acadia19_438.pdf

Kyaw, A. H., Xu, A., Jahn, G., Berg, N., Newnham, C., & Zivkovic, S. (2023). Augmented Reality for High Precision Fabrication of Glue Laminated Timber Beams. *Automation in Construction.* https://doi.org/10.1016/j.autcon.2023.104912

Kyjanek, O., Al Bahar, B., Vasey, L., Wannemacher, B., & Menges, A. (2019). *Implementation of an Augmented Reality AR Workflow for Human Robot Collaboration in Timber Prefabrication.* 36th International Symposium on Automation and Robotics in Construction, Banff, AB, Canada. https://doi.org/10.22260/ISARC2019/0164.

Lee, G. (2022). Code-Bothy: Mixed reality and craft sustainability. *Frontiers of Architectural Research.* https://doi.org/10.1016/j.foar.2022.05.002.

Lok, L., & Bae, J. (2022). Timber De-Standardized 2.0: Mixed Reality Visualizations and User Interface for Processing Irregular Timber. In J. van Ameijde, N. Gardner, H. Hyun, D. Luo, & U. Sheth (Eds.), *Proceedings of the 27th CAADRIA Conference* (pp. 121–130). CAADRIA. https://doi.org/10.52842/conf.caadria.2022.2.121.

Lok, L., Samaniego, A., & Spencer, L. (2021). Timber De-Standardized: A Mixed-Reality Framework for the Assembly of Irregular Tree Log Structures. In B. Farahi, B. Bogosian, J. Scott, J. L. García del Castillo y López, K. Dörfler, J. A. Grant, S. Parasho, & V. A. A. Noel (Eds.), *Proceedings of the 40th Annual Conference of the Association for Computer Aided Design in Architecture* (ACADIA) (pp. 222–231). Association for Computer Aided Design in Architecture (ACADIA). https://doi.org/10.52842/conf.acadia.2021.222

Lok, L., Zivkovic, S., & Spencer, L. (2023). UNLOG: A Deployable, Lightweight, and Bending-Active Timber Construction Method. *Technology|Architecture + Design, 7*(1), 95–108.

Microsoft. (2022). *What is mixed reality? - Mixed Reality*. https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality.

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems, E77-D*(12), 1321–1329.

Rezvani, M., Lei, Z., Rankin, J., & Waugh, L. (2023). Current and Future Trends of Augmented and Mixed Reality Technologies in Construction. In R. Gupta, M. Sun, S. Brzev, M. S. Alam, K. T. W. Ng, J. Li, A. El Damatty, & C. Lim (Eds.), Proceedings of the Canadian Society of Civil Engineering Annual Conference 2022 (pp. 19–39). Springer International Publishing. https://doi.org/10.1007/978-3-031-34593-7_2.

Robert McNeel & Associates. (2022). *Rhino3d* (7.17) [Windows]. Robert McNeel & Associates.

Rutten, D. (2022). *Grasshopper* (1.0.0007) [Windows]. Robert McNeel & Associates. https://www.grasshopper3d.com/.

Skarbez, R., Smith, M., & Whitton, M. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2. https://doi.org/10.3389/frvir.2021.647997.

Song, Y., Koeck, R., & Luo, S. (2021a). [AR]OBOT: The AR-Assisted Robotic Fabrication System for Parametric Architectural Structures. *Blucher Design Proceedings*, 1115–1126. https://doi.org/10.5151/sigradi2021-4.

Song, Y., Koeck, R., & Luo, S. (2021b). Review and analysis of augmented reality (AR) literature for digital fabrication in architecture. *Automation in Construction, 128*, 103762 https://doi.org/10.1016/j.autcon.2021.103762.

Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is Mixed Reality? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300767.

Taylor, J., Bordeaux, L., Cashman, T., Corish, B., Keskin, C., Soto, E., Sweeney, D., Valentin, J., Luff, B., Topalian, A., Wood, E., Khamis, S., Kohli, P., Sharp, T., Izadi, S., Banks, R., Fitzgibbon, A., & Shotton, J. (2016). Efficient and Precise Interactive Hand Tracking through Joint, Continuous Optimization of Pose and Correspondences. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, 2016:35.* https://www.microsoft.com/

en-us/research/publication/efficient-precise-interactive-hand-track
ing-joint-continuous-optimization-pose-correspondences/.
Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., Stühmer,
    J., Cashman, T. J., Tekin, B., Schönberger, J. L., Olszta, P., & Pollefeys,
    M. (2020). HoloLens 2 Research Mode as a Tool for Computer Vision
    Research. ArXiv, Computer Vision and Pattern Recognition. https://doi.
    org/10.48550/arXiv.2008.11239

Publisher's Note