Adaptive Robotic Construction of Wood Frames

Nicholas Cote[1], Daniel Tish[1, 2], Michael Koehle[1], Yotto Koga[1], Sachin Chitta[1]

[1]Autodesk Research, Autodesk, Inc., USA.
[2]Graduate School of Design, Harvard University, USA.

Contributing authors: nick.cote@autodesk.com; dtish@gsd.harvard.edu;

Abstract

Automated robotic construction of wood frames faces significant challenges, particularly in the perception of large studs and maintaining tight assembly tolerances amidst the natural variability and dimensional instability of wood. To address these challenges, we introduce a novel multi-modal, multi-stage perception strategy for adaptive robotic construction, particularly for wood light-frame assembly. Our strategy employs a coarse-to-fine method of perception by integrating deep learning-based stud pose estimation with subsequent stages of pose refinement, combining the flexibility of AI-based approaches with the precision of traditional computer vision techniques. We demonstrate this strategy through experimental validation and construction of two different wall designs, using both low- and high-quality framing lumber, and achieve far better precision than construction industry guidelines suggest for designs of similar dimension.

Keywords: Adaptive Robotics, Robotic Construction, Perception, Machine Learning

1 Introduction

Traditional approaches to wood frame construction and prefabrication depend heavily on manual labor and assembly-line processes, in which parts are moved from station to station and assembled by teams of workers with high-contact tools like mallets, nailers, and tape measures. While well-understood and suitable for high-mix, low-volume production, these approaches can be slow and error-prone. While there is growing exploration into how automation and robotization can accelerate the pace of wood frame construction, the industry faces hurdles in widespread adoption and standardization. Notably, existing commercial solutions optimized for high-volume, low-mix

scenarios and standardized designs face significant challenges for accurate fabrication of highly varied designs amidst the natural variability and dimensional instability of wood [1]. This emphasizes a need for adaptive automated systems that can handle the variation in materials, parts, and processes through the use of perception and modify their actions accordingly. Although robotic solutions in this area are often limited to one-off proofs-of-concept, there is a thriving discussion on adaptivity which can help to address these challenges.

In this work we explore how a multi-modal, multi-stage perception paradigm can enable the automation of wood light-frame assembly tasks. We leverage a strategic coarse-to-fine method of perception, employing consumer-grade sensors, to iteratively refine the precision and adaptability of a robotic construction system specifically geared towards large studs. We demonstrate our approach on a real-world construction project using both low- and high-quality framing lumber, requiring the system to be adaptive and flexible, as shown in Figure 1. We focus on factory-based construction where components or structures are first assembled inside a factory before being installed onsite.

We also employ a basic form of human-robot collaboration that balances automation with human skill. In our setup, a robot is responsible for picking, placing, and holding studs while a construction worker performs the tasks of loading, inspecting, and fastening them. While this introduces several challenges for variability and inaccuracy, it embraces the benefits of an adaptive robotic construction system, discussed later in more detail. This division of tasks also reflects an aspiration to integrate our process into an existing factory-based construction line by replacing a manual wall framing station with a robotized one. Our approach aims to facilitate adoption of automation in such environments without the immediate need for full autonomy, and leverages the dexterity and mobility that skilled human workers bring to the table.

Our work contributes the following:

- A novel multi-stage perception strategy including deep-learning based models trained using synthetic data capable of finding, measuring, and manipulating wooden studs with the precision required for industrial construction.
- Experimental validation of the need for a multi-stage perception strategy and an experimental demonstration of adaptive robotic construction of two wood frame walls using both low and high-quality framing lumber.
- A strategy for incorporating robotic wall framing into existing factory-based construction lines by transitioning from manual to robot-assisted assembly stations.
- Finally, a new benchmark for positional error limits in the robotic construction of wood frames.

The paper is structured as follows: We first review prior work in automation for factory-based construction and highlight challenges and opportunities for automated construction workflows. We then provide an overview of our physical workcell setup and dive into our method of multi-stage perception. We conduct experiments for each stage of perception and demonstrate the precision of our adaptive robotic construction system for two real-world walls, using both low- and high-quality framing lumber. Finally, we discuss insights gleaned from this work and future work in robotic construction.

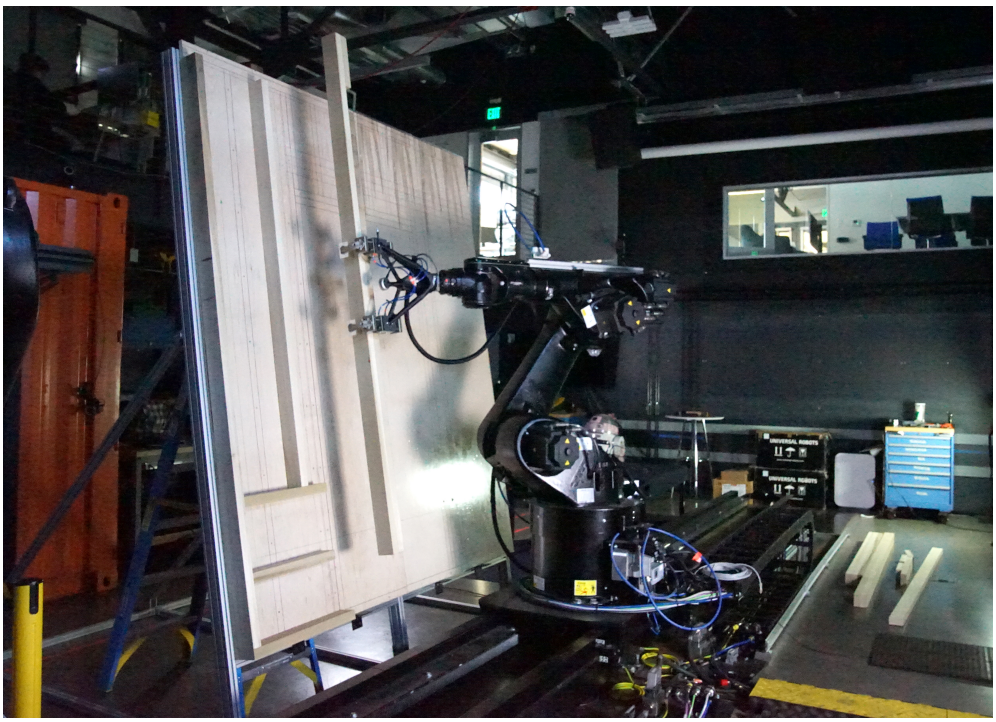

Fig. 1 Mid-assembly snapshot for wall with cantilever detail and high-quality lumber. Shows stud in grippers after measurement stages but before placing stage.

Note that the term *pose* is used throughout this paper in its robotics context, referring to the 6-dimensional combination of position and orientation.

1.1 Related Work

Wood is a common and natural construction material that curls, warps, expands, and deforms. Additionally, there are no set standards for construction tolerances in wood framing, though most sources suggest limiting vertical and horizontal framing error to $< 1/4$ in (6.35 mm) per 8 ft (2.44 m) for manual construction [2–4]. While the manufacturing industry typically expects part dimensions and assembly conditions to be extremely precise (sub-millimeter), the wooden lumber used in construction are imprecise and dimensionally unstable, posing significant challenges for automation. To that end, automated wood frame prefabrication is typically limited to tasks like material preparation (i.e. cutting or shaping parts) and assembly of standardized designs; high-mix scenarios and dexterous tasks still depend primarily on manual labor [1]. Adaptive automation systems, however, could help to navigate the inherent variability and dimensional instability of wood and not only reduce construction error but set a modern standard for construction tolerances in robotic wood framing.

The challenges of robotic wood frame construction are the subjects of ongoing research. Pioneering research at ETH Zurich has demonstrated digital construction

of many parametrically-designed wood frame assemblies, mitigating some of the challenges above by employing high-quality framing lumber [5–8]. Other researchers have attempted to reduce construction error by leveraging motion planning for large, high-aspect ratio framing elements [9, 10], offsite prefabrication and adaptive machining processes to control the geometry of wooden parts [11, 12], and mechanically or force compliant joining techniques [13, 14]. Achim Menges [15, 16] makes a theoretical argument to embrace the material realities of wood and adopt a sensor-driven approach wherein both the available materials and the fabrication process influence the final product. Similar work has focused on scanning unpredictable natural materials during the assembly process [17] and on developing a database of 3D scanned parts [18–20] to optimize their placement in an assembly with respect to design goals and previously placed parts.

In the broader scope of adaptive robotic workflows, computer vision has been implemented to great effect. This is especially true for grasping and bin-picking tasks in warehouse robotics, where an RGBD camera is often used to adaptively pick objects of interest from structured or unstructured environments. Du et al. [21] have conducted a comprehensive review of this area of computer vision-based robotic grasping processes, and identify object localization, pose estimation, and grasp estimation as the core components to achieving this task. Many pose estimation models are trained on synthetically generated data, constructed by simulating different views or positions of 3D CAD models [22, 23]. While many pose estimation models focus on small-scale objects, our research extends those techniques to the larger scale required for construction materials, drawing inspiration from Tish et al. [24] and their strategy for adaptive robotic construction of large façade panels.

In comparison to previous approaches, our approach to robotic wood frame construction focuses on harnessing the latest AI techniques to enable a more flexible workflow capable of dealing with the uncertainty inherent to working with wood as a material and alongside a manual workforce. We use deep-learning perception techniques to locate studs in the workspace, but augment these techniques in a multi-stage strategy with more traditional approaches to enable the higher precision required for our target tasks. Our approach is able to handle wood studs of varying sizes, texture, and shape. We believe that the use of deep-learning based techniques makes our approach more suited to the variability of the wooden studs and the variability in the environment as well. Finally, we demonstrate in an experimental robotic workcell that our approach achieves significantly higher precision for construction of wood frames than industry guidelines suggest, as shown in Figure 7.

2 Workcell Setup

To demonstrate the efficacy of these techniques, we chose the task of constructing full-size wood frame walls using both high- and low- quality framing lumber (see Section 4). We then designed and built a robotic workcell integrating industrial robots, sensors, end-effectors, and auxiliary equipment needed to execute such a task, as shown in Figure 2. We mount a KUKA KR60 industrial robot on a KL1000 linear unit with stations for picking and placing situated in front of it. A picking table, from where

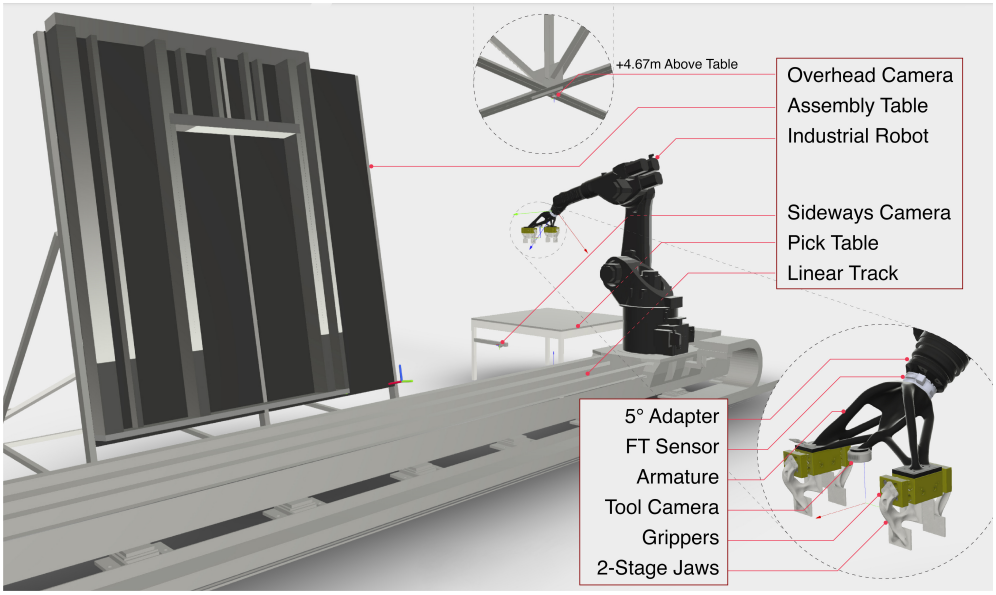

Fig. 2 Workcell layout in simulation with detail of end-effector and wall with doorway condition.

studs are picked up, is horizontal while the placing or assembly table is tilted 80° off horizontal to improve its reachability by the robot. The end-effector is pitched 5° on the flange X-Axis to improve singularity avoidance and two pairs of custom two-stage jaws attached to Schunk PSH grippers are mounted at 45° on the flange Y-Axis. Additionally, an ATI Force/Torque sensor is mounted between the end-effector and the robot flange. We use a closed-source robotics research platform to both simulate and drive our robotic construction process in the real world, creating a digital twin of the workcell, robots, sensors, fixtures, and environmental features [25].

2.1 Camera Configuration

At the core of our multi-stage perception strategy is a comprehensive vision system. An overhead camera, mounted above the picking station, provides an "eye-in-the-sky" view of the picking area. This camera is used for the initial perception stage of stud pose estimation. To capture a 2.74 m (9 ft) stud with 250 mm of buffer on either side, the overhead camera, which had a depth field of view (FOV) of 70° × 55°, must be mounted at least 4.3 m above the table, which is quite far from the part but still within the 9 m working range of the camera. We mounted it to a ceiling truss 4.67 m above the table. Further engineering implications of mounting the camera at this height are discussed in Section 3. To reduce reflections that affect depth readings, the table and surrounding floor area are covered with black fabric. For the subsequent stages of pose refinement, two additional cameras are needed. A tool camera is mounted in an "eye-in-hand" configuration on the end-effector to capture close-up, high-accuracy depth images when grasping. A sideways camera is mounted to the side of the pick table to

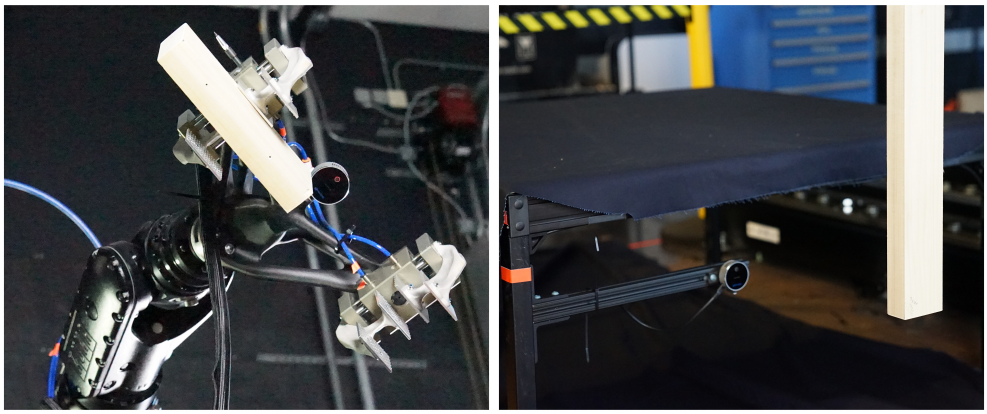

Fig. 3 *Left:* Detail photo showing entire end-effector, tool camera used for centerline estimation stage, and 2-stage dual grippers with shortest stud grasped at its center. *Right:* Sideways camera is shown measuring the in-hand Y-offset of a typical long stud.

measure the offset of a stud held in the grippers. The stud to lens distance for both cameras is maintained at roughly 600 mm for optimal imaging.

The L515 LiDAR camera, shown in Figure 3, was chosen for its high maximum range and reduced depth error over long distances. To discern between the common nominal lumber types of *2x4* (38 x 89 mm), *2x6* (38 x 140 mm), and *2x8* (38 x 184 mm), the depth error at a 4 m distance needed to be less than 22 mm. The L515 depth measurements have an error and standard deviation less than 15 mm at distances up to 9 m. While multiple LiDAR cameras in the same workcell can cause interference issues, we observed few problems due largely to physical line of sight obstructions.

2.2 Studs and Studpacks

In production, studs are manually loaded by a human worker onto the picking table (1 to 4 studs at once) following a known sequence and placing them *near* the center of the table; those parts are, later, picked up and manipulated by a robot. Moreover, the absence of fixtures or automated part feeding systems means that parts can be loaded randomly onto the picking table, posing challenges for traditional, fixed automation systems that depend on precision and consistency. Instead, perception is required to identify and locate parts. This approach is core to the concept of adaptive robotics, wherein sensing and artificial intelligence offer flexibility to a robotic workcell. This approach helps smooth the gap between manual and automated processes, incorporating the skills of an existing workforce at the ends of a robotized construction process. This suggests a path for quicker adoption of automation in industry, promoting a collaborative rather than fully autonomous or fixed approach.

We also group adjacent studs that share a primary axis, as shown in Figure 8, using contact graph analysis, allowing them to be treated as a single component – a studpack – once preassembled and screwed together by a construction worker. While studpacks are a common design feature in wood framing, walls are normally built one

stud at a time, rather than with in groups. In this work, there are several advantages. This strategy decreases the total number of studs and increases the diversity of their geometrical arrangements in the assembly. For workers, this means different parts can be more easily distinguished during loading and grouped parts be prefabricated at a manageable human-scale, in terms of total mass and dimension. For perception, AI models must be trained on a somewhat larger dataset of unique parts. The grouping strategy employed: balances the maximum payload of the robot and its effect on reach, since many studs must be placed near the extremities of the robots Cartesian range; attempts to respect human ergonomics and ensure a manageable carry weight for workers during loading, resulting in part masses that ranged from $1 - 25kg$; and tries to provide a graspable region on each part that fits within the gripper jaws and ensure that the grippers can open after placing a part without colliding with an adjacent one. Additionally, the resulting studs and studpacks vary in length from 0.5 to 2.6 m and arrangement, with L, U, I, and C -shaped cross-sections for studpacks composed of 2-5 individual studs. Per the design, all parts feature only square, or orthographic, end cuts.

2.3 Assembly Process

The assembly process is diagrammed in Figure 4. For each stud loaded into the picking area, the automated system then estimates its pose using the overhead camera and estimates its centerline using the tool camera. The robot then picks up the stud and transports to the sideways camera to measure the in-hand offset, then to the placing station for final placing while monitoring contact; the robot and control system then halt. With the robot now acting as a fixture, a human worker enters the workcell and visually inspects the part for pose errors and provides measured corrections to the robot, repeating as needed. When complete, a worker located safely behind the placing table then fastens the studs to the plywood substrate, driving screws into the studs through labeled, predrilled holes. Note that the construction drawing is printed onto both surfaces of the plywood at 1:1 scale, serving as a visual aid for these workers. Once the worker is finished, they exit the workcell and the robot releases the stud, retracts the end-effector, and returns to the picking table. This process then repeats until all parts are assembled.

3 Multi-Stage Perception

Two challenges in the development of adaptive robotic assembly technologies for construction are the large scale of the parts being assembled and the relatively tight tolerances in the assembly. As noted above, a typical tolerance in the construction of wood frames is roughly 6 mm, which may seem easy to achieve. However, this dimension represents roughly 0.22% of the length of a standard 2.74 m (9 ft) beam. This very small ratio of tolerance to overall length is a challenge for many perception technologies, but especially for vision systems, which can be affected by sensor noise at large distances and restrictions on image resolution. Moreover, capturing such large studs requires cameras to be positioned further away, reducing pixel accuracy and complicating the assembly process.

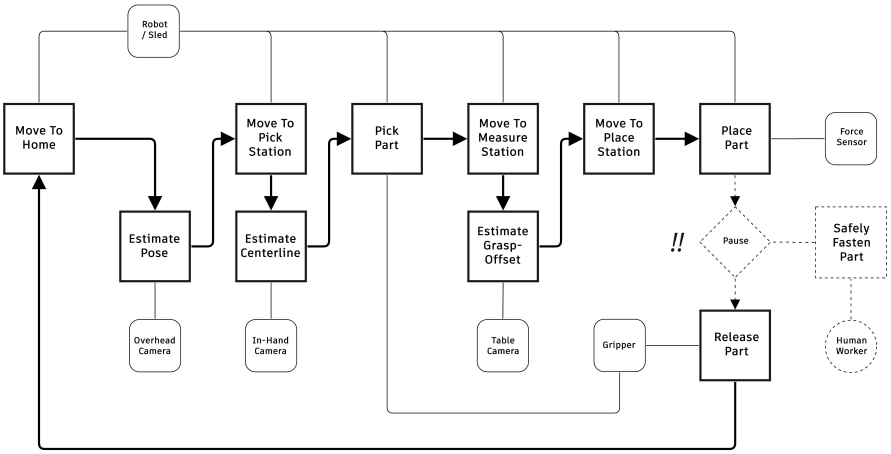

Fig. 4 Robotic construction process starting from top left. Shows robot actions, such as *Move To Home*, perception stages, such as *Estimate Pose*, and devices in the workcell, such as *Robot* and *Overhead Camera*.

3.1 Coarse-to-Fine

To tackle these challenges, we developed a multi-stage perception method which progressively increases resolution along critical dimensions and refines pose estimations, as shown in Figure 5. This "coarse-to-fine" strategy has several advantages. It allows the system to balance the flexibility of the AI-driven first stage with the accuracy of traditional methods in subsequent stages. Thus, each stage is highly-specialized in one aspect of the pose estimation process, making the overall process robust to the variability of wood and able to achieve the high-precision required for robotic construction tasks. This nuanced approach of trading off flexibility and precision at different stages addresses the unique challenges of robotic wood frame construction and leverages the strengths of both AI-driven and traditional measurement techniques.

First, we employ a deep learning model for *initial pose estimation* using the overhead camera. We then refine the X, Y, and Rz components of the initial pose by *measuring the centerline* of the stud using the tool camera before it is grasped and *measuring the in-hand offset* of the stud using the sideways camera after it is grasped. Given that the stud lies flat on a table, the Z component of its position relative to the table is equal to its depth (assuming it's accurately identified), while both the Rx and Ry components of its rotation are nearly zero and may be ignored. Once a stud is picked accurately, we can generally assume it will be placed accurately, however, a final stage of *force-based contact detection* was added to mitigate calibration error and stack-up of errors caused by dimensional variation of studs and assembly tolerances.

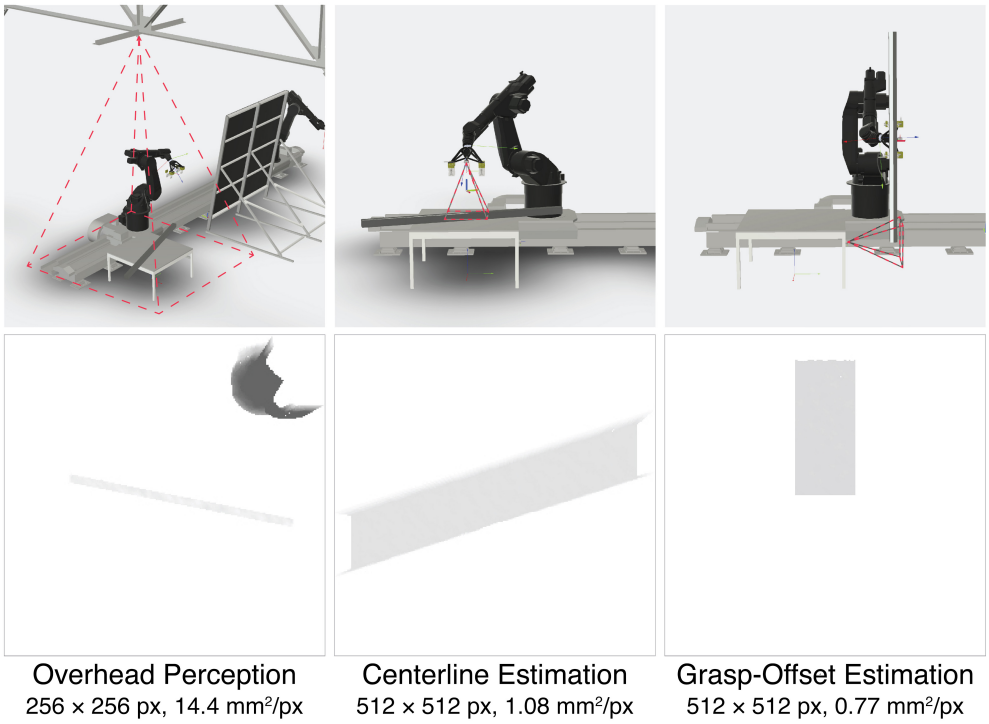

Fig. 5 *Left to right*: Overhead Perception, Centerline Estimation, and Offset Estimation showing pose of camera relative to stud (top) and typical thresholded depth image (bottom) for each stage.

3.2 Initial Pose Estimation

The Initial Pose Estimation stage locates parts on the table using a depth image captured from a ceiling-mounted camera. This image shows the entire stud and is used to approximate its pose prior to further stages of refinement.

The adaptive assembly process begins with a deep learning-based pose estimation algorithm, described in Koga et al. [25]. This algorithm uses a DenseNet architecture to regress a depth image taken from the overhead camera into a 6-DOF pose for the stud in the world frame, a technique that has demonstrated effectiveness in a variety of AI-driven perception tasks [26]. Note that we've adapted this implementation to return a single grasp pose at the top center of the stud, rather than infer multiple possible grasp poses from a single image. The model is trained on simulated data, generated by randomly placing the CAD model of the stud in the picking area, capturing a depth image using the overhead camera, converting this to a depth-thresholded orthographic point cloud to remove background objects (such as the table) and isolate the stud, and then by computing a grasp proposal for the stud therein. This proposal coincides with the 3D center and orientation of the point cloud, to which we add half the depth of the stud to get its top for grasping downstream. As described in Koga et al., this training step is performed 600,000 times for the set of parts in a given assembly and

takes roughly 12 hours on a Nvidia V100 GPU. A typical point cloud derived from this depth image of a stud is shown in Figure 6. The simulation incorporates Perlin noise to bridge the gap between synthetic and real-world data, ensuring the model's robustness in real-world applications.

However, the perception model architecture is constrained to 256 x 256 pixel depth images, which, at the time, could not easily be increased. In bin-picking or small-scale assembly tasks, this resolution normally provides an acceptable degree of accuracy balanced with training efficiency. However, at the distance of our overhead camera, described in Section 2, the accuracy of the pose estimates is considerably low. Each pixel in the depth image represents a 14.4 mm square on the table, accurately locating an edge or a corner becomes unfeasible, as this dimension is greater than our assembly tolerances and industry guidelines.

Acknowledging these limitations, we opted for a multi-stage approach that aligns with the realities of current factory environments, where implementing advanced, high-resolution systems might not be immediately feasible due to cost, complexity, or integration challenges. Although this approach may seem less sophisticated than a single AI-driven stage, it allows for the use of readily deployable AI models supplemented by traditional perception techniques. Moreover, this approach could facilitate the adoption of automation in real-world manufacturing environments and enable incremental improvements alongside advances in AI and sensor technology. With the availability of higher resolution images, such as 512 x 512 pixels, or industrial-grade sensors, it may be possible for pose estimates provided by this stage to be accurate enough that subsequent stages of refinement can be avoided entirely.

In our experiments we found that subsequent stages of perception provided acceptable results as long as the initial pose estimate was within 150 mm for X and Y and within 15° for Rz relative to the center of the table. The pose for studs positioned near the edges of the table proved more difficult to estimate using the chosen method, suggesting an optimal region of interest within the camera's view (and smaller than we initially expected). Hence, we defined the loading area carefully so that studs would start within this region. Then, when we encountered an obviously incorrect estimate, we simply manually relocated the stud within this region and tried again.

3.3 Centerline Estimation

The Centerline Estimation stage ensures accurate picking by capturing a depth image with the tool camera from above the picking table origin or previously pose. This image shows an uninterrupted section of the stud's grasping area, allowing for the calculation of the stud's centerline for precise picking.

Because studs are generally much larger than the captured image, they appear to clip at the edges. The center and rotation estimation are achieved through a two-stage Principal Component Analysis (PCA) of the point cloud returned from the depth image from the tool camera. The first stage is used to transform the point cloud into an axis-aligned 3D space and to throw out any points that exceed our variance threshold. The second stage is performed on the transformed point cloud derived from the first analysis. The first eigenvector from this analysis gives the vector of the centerline of the stud, with which we can refine Rz. Using the mean of the point cloud and the

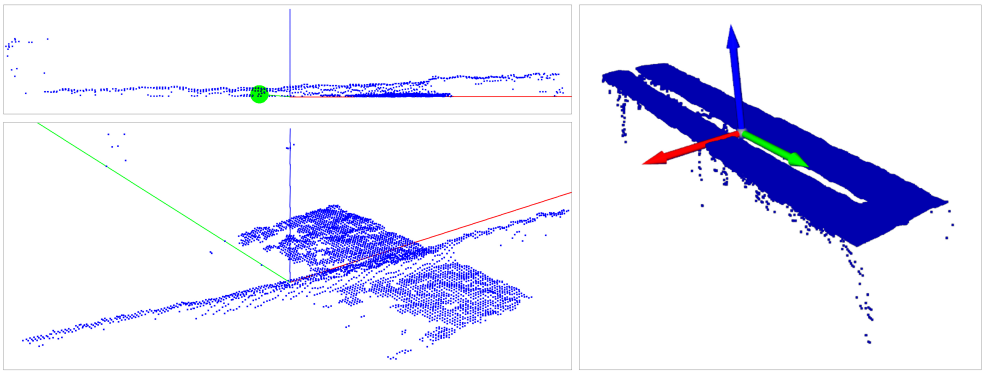

Fig. 6 *Left, top and bottom*: Example point cloud for a typical stud, converted from a depth image taken by the overhead camera during Initial Pose Estimation stage. *Right*: Example point cloud for a typical studpack showing estimated grasping pose after Centerline Estimation stage.

known position of the tool camera, with which we can refine the X component of the stud's center. The result of this process for a typical studpack is shown in Figure 6.

This two-stage process is useful for removing artifacts from the camera frame in the point cloud. Once transformed into the centered and axis-aligned space from the first analysis, a box clipping method trims the parallelogram-shaped point cloud into a rectangle to maintain the accuracy of the centerline vector estimation. The need for this trimming process is mitigated by aligning the camera's axes with the stud's axes, estimated using the ML process. This process of alignment is not possible in every circumstance, however. In the rare case that a notch or other feature produces a non-continuous width in the camera view, a similar method is used to estimate the center and rotation of the stud from its rectangular bounds, rather than relying on the PCA analysis. In this analysis, the centerline is assumed to be parallel to the long edges and the center estimation is the midpoint of the diagonal.

For short studs (< 600 mm) that can easily fit into the tool camera frame, a visual servoing method is used to ensure that the entire stud is visible before starting the refinement process. The method checks whether an object exists at the edge of the depth image and, if so, steers the robot above the previous position and in the direction of the violated edge. The estimated center and on-table rotation are used as the target point for the picking operation. While an accurate Z coordinate of the center point is typically returned by the function, we use the known heights of the pick station table and the stud, extracted from their CAD models, instead to prevent collisions with the picking table.

3.4 In-Hand Offset Estimation

The In-Hand Offset Estimation stage ensures accurate placement by using a sideways-mounted camera to capture a a depth image of one end of the stud. Using this image, we can measure the distance of the end of the stud to the grasp location, ensuring the stud is placed accurately based on where it was grasped.

With the stud firmly grasped, the Y component of the pose can be refined by measuring the offset along the longitudinal axis of the stud. One end of the stud is gradually passed in front of the sideways camera and stopped when its bottom edge reaches the horizontal centerline of the image, using traditional depth thresholding and edge detection techniques to do this. From this position, we subtract the Z component of the camera position from that of the Tool Center Point (TCP), thus measuring the distance from the TCP to the lower edge of the stud. Ideally, because the stud is grasped at its geometrical center, this distance is equal to half the total length of the stud. So, to calculate the actual in-hand offset from the ideal, we simply subtract half the length of the stud from our measurement. This offset is then added to Y component of the grasp transformation matrix, which locates the grasp in the coordinate system of the stud, and enables accurate placing downstream.

Since we have carefully measured and cut studs to their designed lengths ahead of time, we need only perform this measurement once. In a scenario where cut lengths are not guaranteed to be accurate, both ends of the part must be measured either simultaneously, such as with a second camera, or sequentially, such as by flipping the part and repeating this process.

3.5 Force-based Contact Detection

The Force-based Contact Detection stage further ensures accurate placement against an imperfect work surface by monitoring contact forces using a Force-Torque sensor on the end-effector. This ensures the part securely abuts the table or adjacent parts before being fastened.

Before placing a stud, we assume that its pose in the gripper is well-known, and that both the end-effector and place-table have been well-calibrated prior to program start. To compensate for any remaining translation and rotational error (< 3 mm, $0.1°$), we implement contact monitoring. During placing, the stud is translated along a series of approach vectors towards its final pose until it either reaches that pose or abuts the table or an adjacent stud, at which point a force-monitoring algorithm identifies that contact has been made and halts the motion command. After the robot has stopped moving, the stud is considered "placed".

To compensate for both sensor bias and gravity and to isolate contact forces on each approach, we first take a contact-less reading from the sensor then subtract it from subsequent readings. To filter noise and identify an upward trend consistent with contact, we use an exponentially-weighted moving average for each force-torque component and their magnitudes. Taking advantage of some material compliance in the end-effector, we also allow the robot to press the stud against the table and "flatten" any rotational errors caused by deflection in the end-effector or due to poor calibration. We also subtract the actual position of the stud from its design position to calculate the accumulated global drift of the assembly.

3.6 Final Assurances

To verify the precision of stud placement after the previous stages, a human worker conducts a visual inspection of the stud before the robot releases it, as noted in Section

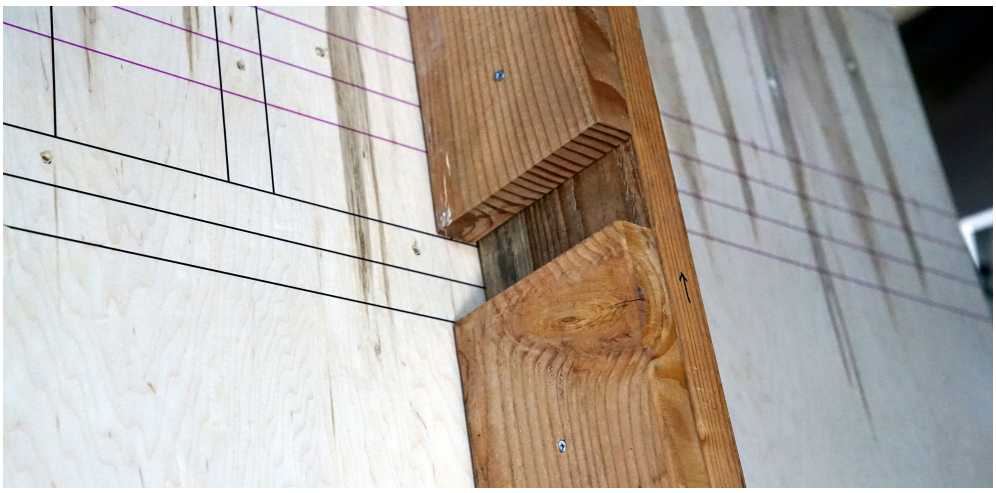

Fig. 7 Low-grade studpack after being placed by the robot, showcasing high-precision of our method. Design drawings printed on wall for reference.

2. In the event of a minor error, the worker measures it by hand and inputs the necessary Cartesian offsets into the control terminal, which the robot applies after they exit the workcell. This corrective loop continues as needed. For severe errors, such as a collision or incorrect part orientations, the worker inputs a reset command, prompting the robot to return to the pick table, release the part, and restart the entire process. While we intend to automate this stage in future work, it was largely unnecessary in practice due to the high accuracy of previous stages.

4 Experiments

We conducted experiments in simulation and reality to evaluate the efficacy of our method. We assess each perception stage by having the robot pick up a stud and place it in a known, albeit randomized, location 10 times, running its estimation function 10 times. We then construct two light-frame walls to demonstrate the practical application of this technology.

4.1 Stage Evaluation

Two conditions were tested for the centerline estimation experiment: one where the overhead perception stage provided a valid result and the camera can be "aligned" to the stud; and another where it failed and the camera must be "centered" over the table. This allows us to capture the shortest distance between the actual centerline of the stud and the estimated center point, in addition to any errors in on-table rotation (Rz). For the in-hand offset estimation, the stud was grasped from known locations along the length of the studpack and passed in front of the fixed camera as previously described.

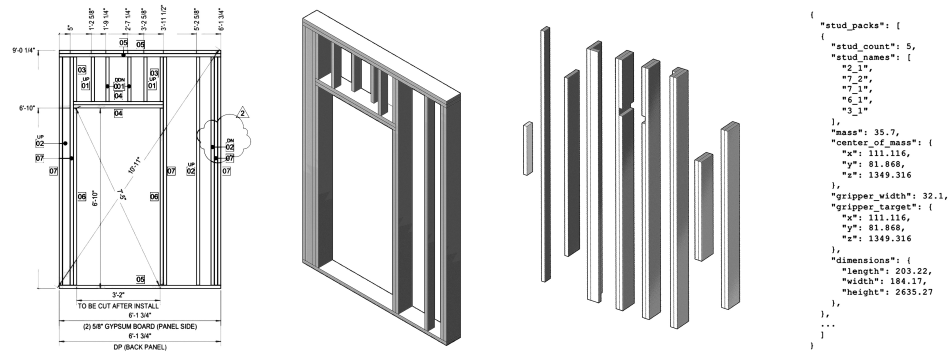

Fig. 8 *From left:* Based on 2D construction drawings for a given wall (first), we create a 3D model of the entire assembly (second) and, then, identify unique studs and studpacks (third). Note that several unique parts with varying geometries are shown, some of which are used more than once in the design. Data needed for robotic assembly, such as part pose and name, is then saved to a JSON (fourth).

These experiments highlight the necessity and effectiveness of a multi-stage approach, as in Table 4.1. In the initial pose estimation stage, the average error was 2.15 mm and 11.08 mm along the X and Y axes. Additionally, results showed considerable variance, with 14.01 and 76.31 mm variance for X and Y axes and 0.05° for Rz. Through the subsequent centerline and offset estimation refinements, we reduce these errors to < 1 mm in both positional dimensions and achieving nearly perfect rotational estimations. The precision of the results also improves dramatically, with a full variance of results in X and Y of 4.17 mm and 2.73 mm, respectively, and a very low standard deviation across the board. Note that these figures are well-below the established error limits suggested for the construction of wood frames.

When loading studs onto the pick table, their shortest and longest edges were aligned perpendicular (±15°) to the world X and Y axes. This setup limited the number of pixels available for Y-estimation during overhead perception, as shown in Figure 9. Interestingly, the in-hand estimation error is 43% that of the centerline estimate despite its pixel resolution being slightly larger.

Experiment	Notes	Axis	Units	Average	STD	Variance
Pose Estimation	-	X	mm	2.15	1.72	14.01
Pose Estimation	-	Y	mm	11.08	9.01	76.31
Pose Estimation	-	Rz	deg	0.012	0.009	0.05
Centerline Refinement	*Aligned to origin*	X	mm	1.45	0.71	4.61
Centerline Refinement	*Aligned to stud*	X	mm	0.93	0.70	4.17
Centerline Refinement	*Aligned to origin*	Rz	deg	0.005	0.008	0.034
Centerline Refinement	*Aligned to stud*	Rz	deg	0.0005	0.0003	0.002
In-Hand Refinement	-	Y	mm	0.40	0.33	2.73

Table 1 Dimensional error measured in experiments.

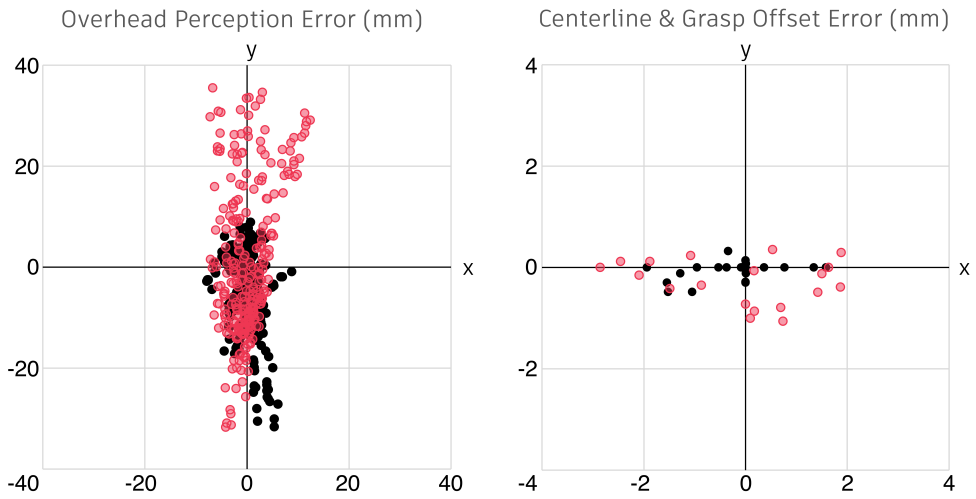

Fig. 9 Estimated position of stud (points) relative to ground-truth position of stud (origin), showing spatial trends in error for Overhead Estimation stage (left) and combined Centerline and In-Hand Estimation stages (right). Simulated estimates in black, real-world estimates in red. *Right:* X-Axis: Centerline Error, Y-Axis: In-Hand Error

4.2 Wall Construction

We built two types of walls: a 1.83 x 2.74 m (6 x 9 ft) wall with a doorway header condition and a 2.74 x 2.74 m (9 x 9 ft) wall with a cantilevered upper section. These walls are typical in a factory-based modular construction project and of manageable dimensions for our robots and workcell. Importantly, they were also chosen because they include a modest range of stud configurations and geometries typical in wood frame construction, as shown in Figure 8. The first wall was constructed using low-grade framing lumber which featured knots, curling, and warping. The second wall was constructed using clean, kiln-dried poplar with excellent dimensional stability. For both walls, additional tolerance of 1 mm was created at notched joints to ease tight-fit assembly conditions for warped and imperfect studs.

The framing models for these walls were extracted from a larger Building Information Model (BIM) and come from a partnership with a local construction company on a real-world project. We extracted the physical properties (i.e. mass, center of mass), geometrical properties (i.e. dimensions, center of geometry), and target pose from the framing model and define the assembly sequence, approach vectors, and grasping pose for each stud. Finally, we export this data to a JSON file, export a mesh of each stud from a common origin, and then fabricate the physical studs for assembly experiments. From the CAD model we also generate a mapping of screw locations for each stud and, then, drill thru holes in the plywood substrate to facilitate a more intuitive and physically easier fastening process.

Images of the two completed wall frames, with detailed mid-assembly images highlighting a tight insertion task and an eccentric gripping condition are shown in Figure

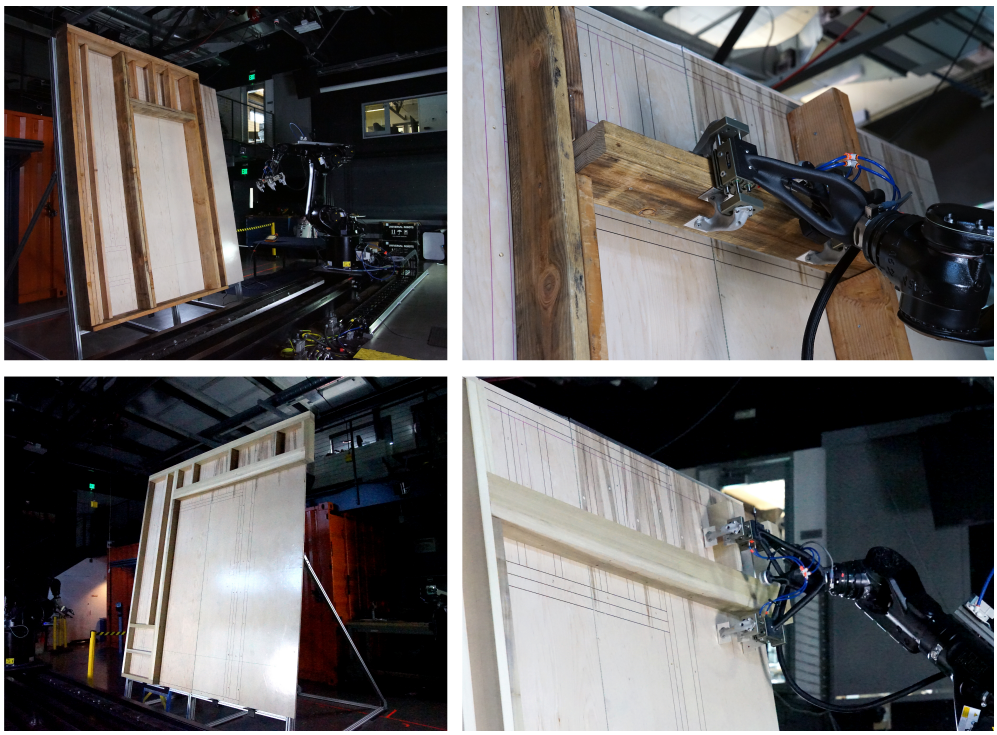

Fig. 10 *Top left:* Doorway wall with low-quality wood. *Bottom left:* Cantilevered wall with high-quality wood. *Top right:* Tight fit insertion task. *Bottom right:* Complex grasp-spanning task.

10. All studs were successfully placed as designed, and our method consistency delivered results below the 6 mm tolerance given by the force-based contact detection, ensuring there were no unwanted gaps in the assembly or misaligned parts. Although we had planned for a human worker to inspect the stud and suggest pose corrections after placement, parts were place precisely enough that those corrections were not needed.

5 Conclusions

This work demonstrates a novel multi-modal, multi-stage perception strategy for adaptive robotic construction of wood frames that is precise, robust to the inherent variability of framing lumber and manual work, and which can handle reasonably diverse stud configurations. We provide a viable model for incorporating the skills, dexterity, and mobility of human workers in a factory-based construction environment that's transitioning from manual processes to ones that are robot-assisted or fully automated. We show that balancing the flexibility of AI-driven pose estimation techniques and the precision of traditional vision-based measurement techniques can help address the unique challenges of robotic construction. Moreover, we show that higher accuracy than is typically suggested for manual wood frame construction is achievable,

even when using inexpensive sensors and low-grade framing materials. Thus, we offer our results as a sub-millimeter benchmark for limiting error in robotic construction of wood frames. In conclusion, our work represents a significant stride towards overcoming the barriers to the adoption of robotic solutions in wood frame construction, and we are eager to see how this strategy can be scaled to meet the growing demands of this industry.

6 Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] E. Lachance, N. Lehoux and P. Blanchet, "Automated and robotized processes in the timber-frame prefabrication construction industry: A state of the art," 2022 IEEE 6th International Conference on Logistics Operations Management (GOL), Strasbourg, France, 2022, pp. 1-10, doi: 10.1109/GOL53975.2022.9820541.

[2] National Association of Home Builders (NAHB). Residential Construction Performance Guidelines, Contractor Reference. BUILDERBOOKS, 2022.

[3] K., D., and Ballast. Handbook of Construction Tolerances, 2nd Edition. John Wiley & Sons, 2007.

[4] RS Means. Residential & Light Commercial Construction Standards. RS Means, 2008.

[5] P. Eversmann, F. Gramazio, and M. Kohler, “Robotic prefabrication of timber structures: towards automated large-scale spatial assembly,” Constr Robot, vol. 1, no. 1, pp. 49–60, Dec. 2017, doi: 10.1007/s41693-017-0006-2.

[6] A. Thoma, A. Adel, M. Helmreich, T. Wehrle, F. Gramazio, and M. Kohler, “Robotic Fabrication of Bespoke Timber Frame Modules,” in Robotic Fabrication in Architecture, Art and Design 2018, Cham, 2019, pp. 447–458. doi: 10.1007/978-3-319-92294-2_34.

[7] J. Willmann, M. Knauss, T. Bonwetsch, A. Apolinarska, F. Gramazio, and M. Kohler. “Robotic Timber Construction: Expanding Additive Fabrication to New Dimensions.” Automation in Construction 61, 2015 , pp. 16-23. doi: https://doi.org/10.1016/j.autcon.2015.09.011

[8] P. Y. Leung, A. Apolinarska, D. Tanadini, F. Gramazio, and M. Kohler, “Automatic Assembly of Jointed Timber Structure using Distributed Robotic Clamps,” in PROJECTIONS - Proceedings of the 26th CAADRIA Conference, Hong Kong, Mar. 2021, vol. 1, pp. 583–592. doi: 10.52842/conf.caadria.2021.1.583.

[9] I. A. Sucan, M. Moll, and L. E. Kavraki, “The Open Motion Planning Library,” IEEE Robotics & Automation Magazine, vol. 19, no. 4, pp. 72–82, Dec. 2012, doi: 10.1109/MRA.2012.2205651.

[10] A. Gandia, S. Parascho, R. Rust, G. Casas, F. Gramazio, and M. Kohler, “Towards Automatic Path Planning for Robotically Assembled Spatial Structures,” in Robotic Fabrication in Architecture, Art and Design 2018, Cham, 2019, pp. 59–73. doi: 10.1007/978-3-319-92294-2_5.

[11] H. J. Wagner, M. Alvarez, A. Groenewolt, and A. Menges, “Towards digital automation flexibility in large-scale timber construction: integrative robotic prefabrication and co-design of the BUGA Wood Pavilion,”Constr Robot, vol. 4, no. 3, pp. 187–204, Dec. 2020, doi:10.1007/s41693-020-00038-5.

[12] J. Pedersen, A. Søndergaard, and D. Reinhardt, “Hand-drawn digital fabrication: calibrating a visual communication method for robotic on-site fabrication,” Constr Robot, vol. 5, no. 2, pp. 159–173, Jun. 2021, doi: 10.1007/s41693-020-00049-2.

[13] A. Apolinarska, M. Pacher, H. Li, N. Cote, R. Pastrana, F. Gramazio, and M. Kohler, "Robotic assembly of timber joints using reinforcement learning". Automation in Construction, vol. 125, 2021. doi: https://doi.org/10.1016/j.autcon.2021.103569

[14] S. Lewis, N. King, G. Fagerstrom, C. Luo, and N. Cote, “Toward an Automated Robotic Fabrication Workflow for Structurally Optimized Multispecies Wooden Network Shells.” Proceedings of IASS Annual Symposia, Vol. 2018, No. 20. International Association for Shell and Spatial Structures (IASS), 2018.

[15] A. Menges, “The New Cyber-Physical Making in Architecture: Computational Construction,” Architectural Design, vol. 85, no. 5, pp. 28–33, 2015, doi: https://doi.org/10.1002/ad.1950.

[16] G. Brugnaro, E. Baharlou, L. Vasey, and A. Menges, “Robotic Softness: An Adaptive Robotic Fabrication Process for Woven Structures,” Ann Arbor (Michigan), USA, 2016, pp. 154–163. doi: 10.52842/conf.acadia.2016.154.

[17] L. Vasey, T. Grossman, H. Kerrick, D. Nagy. "The hive: a human and robot collaborative building process". ACM SIGGRAPH 2016 Talks, 1-2, 2016. doi: 10.1145/2897839.2927404.

[18] G. Furusho, Y. Nakamura, and G. Hirasawa, “Raw Wood Fabrication with Computer Vision,” in Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC), Dubai, UAE, Nov. 2021, pp. 741–746. doi: 10.22260/ISARC2021/0100.

[19] K. Wu and A. Kilian, “Designing Natural Wood Log Structures with Stochastic Assembly and Deep Learning,” in Robotic Fabrication in Architecture, Art and

Design 2018, Cham, 2019, pp. 16–30. doi: 10.1007/978-3-319-92294-2_2.

[20] Y. Qi et al., "Augmented Accuracy: A human-machine integrated adaptive fabrication workflow for bamboo construction utilizing computer vision," in Towards a new, configurable architecture - Proceedings of the 39th eCAADe Conference, Novi Sad, Serbia, Aug. 2021, pp. 345–354. doi: 10.52842/conf.ecaade.2021.1.345.

[21] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review," Artif Intell Rev, vol. 54, no. 3, pp. 1677–1734, Mar. 2021, doi: 10.1007/s10462-020-09888-5.

[22] Y. Litvak, A. Biess, and A. Bar-Hillel, "Learning Pose Estimation for High-Precision Robotic Assembly Using Simulated Depth Images." arXiv, Mar. 23, 2019. doi: 10.48550/arXiv.1809.10699.

[23] J. W. Ma, T. Czerniawski, and F. Leite, "Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds," Automation in Construction, vol. 113, p. 103144, May 2020, doi: 10.1016/j.autcon.2020.103144.

[24] D. Tish, N. King, and N. Cote, "Highly accessible platform technologies for vision-guided, closed-loop robotic assembly of unitized enclosure systems," Constr Robot, vol. 4, no. 1, pp. 19–29, Jun. 2020, doi: 10.1007/s41693-020-00030-z.

[25] Y. Koga, H. Kerrick, and S. Chitta, "On CAD Informed Adaptive Robotic Assembly." arXiv, Aug. 02, 2022. doi: 10.48550/arXiv.2208.01773.

[26] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation." arXiv, Oct. 31, 2017. doi:10.48550/arXiv.1611.09326.

[27] "KR 30, 60-3; KR 30 L16-2 with KR C4 Operating Instructions, Version 05." KUKA Roboter GmbH, Jan. 25, 2017. Accessed: Mar. 26, 2021. [Online]. Available:https://www.kuka.com/en-us/services/downloads