Material-Aware Contextual Intention Inference for Adaptive Human–Robot Collaborative Timber Assembly

Yiqing Wang[1], Zhengyi Chen[2] and Xingyu Tao[1]

[1]BRE, Hong Kong Polytechnic University, Hong Kong, China
[2] School of Civil Engineering, Southeast University, China
wangyiqing4214@gmail.com, zhengyi.chen@seu.edu.cn, xingyu.tao@polyu.edu.hk

Abstract

Designing adaptive human–robot collaboration (HRC) systems for timber construction is challenging due to the difficulty of encoding tacit construction knowledge and material variability into computational models. This paper presents a material-aware probabilistic framework for light-frame timber assembly. A lightweight timber construction ontology is extracted from construction manuals and empirical observations to formalize component relationships and assembly logic. Based on this representation, a temporal hidden semi-Markov model (HSMM) is trained from fully supervised state traces from simulation results, while a vision-based macro-stage prior is learned from on-site material stock to provide contextual anchoring. The two models are integrated in a material-aware contextual task prediction algorithm, where human accept/reject feedback is treated as external evidence to update HSMM beliefs and state durations. The framework is implemented in an HSMM-based material-aware HRC workplace allowing interactive handover of timber assembly components. Experimental results show that material-aware priors substantially improve interaction performance, doubling the accept rate (0.25 → 0.58) and reducing rejection loops, with consistent gains observed in real HRC experiments.

Keywords –

Material-Awareness; Intention Inference; HSMM; Timber Construction; Human-Robot Collaboration

1 Introduction

Timber construction remains highly labor-intensive and difficult to automate due to the complexity of operations and the large number of variables present on construction sites (Herzog et al., 2012), including diverse timber elements, evolving material conditions, and dynamic on-site availability. Human workers therefore rely heavily on empirical knowledge to determine efficient and feasible assembly sequences, many of which are not explicitly defined during the design phase but instead emerge through experience-informed, construction-specific reasoning. As a result, existing robotic systems are often inefficient or overly predetermined, limiting their ability to adapt to changing material conditions, on-site stock, and locally optimal assembly sequences that are difficult to predefine. Enabling robotic systems to infer human construction intentions and contextual decision-making is thus essential for effective human–robot collaboration and the transfer of tacit construction knowledge into adaptive automated processes.

Early work in intention inference relied on rule-based task models or predefined state machines, which are effective in structured manufacturing environments but require explicit task definitions and fixed assembly sequences (Adams et al., 1996; Wang et al., 2018). To improve flexibility, subsequent research has explored learning-based and multimodal intention inference. Vision-based methods integrate human motion recognition, object detection, and contextual cues such as object affordance to infer task intent (Cramer et al., 2018; Zhang et al., 2022; Yu et al., 2023). Related work formulates collaborative assembly as a graph-based decision problem and applies reinforcement learning under predefined task structures (Zhang et al., 2022). In parallel, physical interaction–driven approaches infer intention from force, impedance, or co-manipulation dynamics to support close-proximity collaboration, particularly in handover and shared manipulation scenarios (Peternel et al., 2016; Wang et al., 2022; Yu et al., 2023). While effective for local interaction safety and responsiveness, these methods largely focus on motion-level or short-horizon physical intent.

More recent studies investigate temporal and probabilistic formulations for real-time and predictive intention inference, reasoning over task or interaction states under uncertainty (Gomez et al., 2021; Nottensteiner et al., 2021; Liu et al., 2023). However, these approaches typically assume predefined state spaces and infer intent primarily from kinematic observations, with limited integration of material conditions or dynamic resource availability.

Despite these advances, several critical gaps remain in the literature when applied to HRC timber construction scenarios. First, task inference in timber assembly is inherently challenging due to high levels of uncertainty arising from evolving material stock, incomplete observations, and context-dependent decision-making. Many existing approaches rely on predetermined task structures or predefined state transitions, which limit their ability to adapt to the non-deterministic and emergent nature of construction processes. Such methods often fail to capture the inherent uncertainty and flexibility observed in real-world timber assembly. While perception-driven and multimodal approaches improve responsiveness, they are often computationally intensive and rely heavily on high-dimensional sensory inputs and extensive training. This makes them less suitable for complex and dynamic construction environments, where scalability across different building types and material configurations is required.

To address this gap, this paper proposes a material-aware, lightweight, and scalable task inference framework based on a hidden semi-Markov model, enabling contextual intention prediction under evolving material conditions and limited observation for adaptive human–robot collaborative timber assembly.

To achieve this goal, three objectives are defined:

• an ontology for light-frame timber construction was developed enabling structured yet non-prescriptive task representation for adaptive assembly;

• a material-aware HSMM-based model that integrates contextual macro-stage information derived from material distribution, allowing inference beyond predefined task sequences;

• to implement an adaptive human–robot collaboration workflow that leverages material context and human feedback for interactive task prediction under uncertainty

2 Methodology

The proposed methodology integrates construction knowledge, material-aware perception, and probabilistic temporal inference to enable adaptive human–robot collaboration in light-frame timber assembly (Figure 1).

First, empirical knowledge from timber construction manuals is formalized into an assembly ontology defining seven micro-level assembly states grouped into three macro stages (slab, wall, roof). Minimal hard constraints encode essential structural dependencies, preventing physically invalid transitions without prescribing a fixed assembly sequence.

Second, ten representative timber assemblies are simulated in a Rhino/Grasshopper/GhPython environment to generate ordered state traces for training a hidden semi-Markov model (HSMM). The HSMM learns state transition probabilities and duration distributions, while a vision-based module extracts material stock features from on-site layouts to estimate a probabilistic macro-stage prior conditioned on available components.

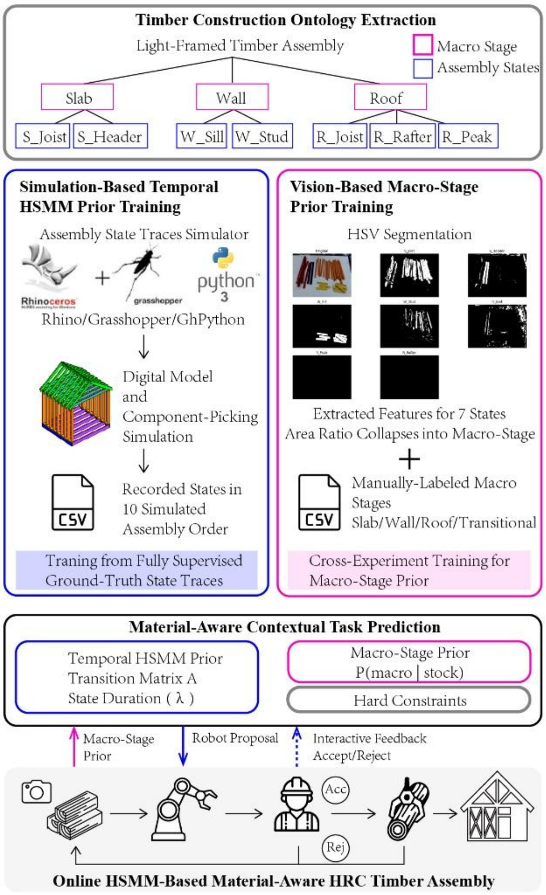

Figure 1. Material-Aware Task Prediction Framework for HRC Timber Construction

Finally, during collaboration, the HSMM maintains a belief over assembly states and proposes candidate next actions. Sparse human accept/reject feedback is incorporated to update the belief and suppress repeated incorrect proposals, while the material-derived macro-stage prior provides contextual anchoring for state inference. This framework enables intention-aware, material-conditioned task prediction under uncertainty while remaining responsive to human judgment.

2.1 Timber Construction Ontology Extraction

Construction knowledge is first derived from a timber construction manual (Herzog et al., 2012) and structured through a semantic representation layer that captures construction methods, material types, construction phases, assembly grouping and order, tools, and operations. This prior knowledge including material types, assembly order and interrelations is subsequently simplified and refined for simulation and physical experiments in which material stock observations and assembly sequences are explicitly modeled.

The resulting empirical patterns are extracted as structured knowledge and formalized into ontology concepts and rule-based constraints (Figure 2). In this study, a simplified demonstrative task based on a typical light-framed timber house structure is used to instantiate the proposed representation (Figure 3).

Figure 2. Timber Structure Ontology

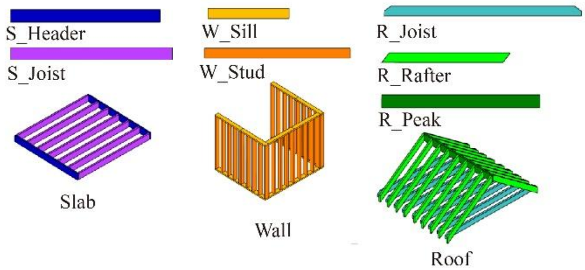

Figure 3. Modeled Macro Stages and Micro States

2.2 Simulation-Based Temporal HSMM Prior Training

A custom simulator was developed in Rhino/Grasshopper with GhPython to generate sequential task records for representative light-frame timber structures (Figure 4). The simulator provides an interactive interface in which a human operator selects the timber element to be assembled at each step, emulating construction sequencing decisions. Each selection, reflecting both global and local assembly order, is recorded with a timestamp and material label and exported as a structured CSV log.

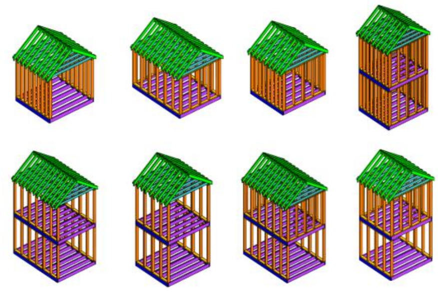

Figure 4. Modeled Timber Structure Types

Using this interface, ten simulated assembly sequences were generated, including sequencing variations across two representative structure types selected from eight modeled timber house configurations. These logs serve as fully supervised ground-truth state traces for training the temporal hidden semi-Markov model (HSMM). State transition probabilities (Figure 5) and duration priors (Figure 6) are learned directly from the recorded assembly sequences, forming the temporal basis for online inference. The learned model captures three dominant cyclic patterns consistent with light-frame timber construction: slab assembly with extended joist runs and occasional headers; wall assembly with strong stud–sill alternation; and roof assembly with brief joist transitions followed by long rafter sequences and single-step peak completion. Because our training traces only contain ground-truth accepted states and do not include a sensor observation stream, we estimate only the HSMM temporal parameters (A, λ).

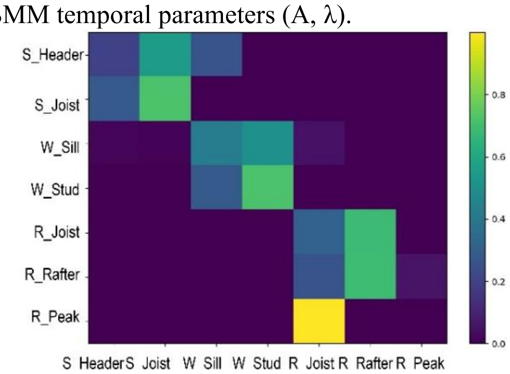

Figure 5. HSMM State Transition Matrix A

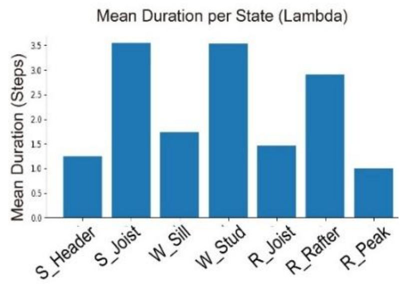

Figure 6. HSMM Mean Duration Results

2.3 Vision-Based Macro-Stage Prior Training

Three physical experiments were recorded, and top-view images of on-site material stock were used to extract features representing material availability over time. Because this study focuses on intention inference rather than visual perception, assembly components corresponding to the seven micro-states were color-coded to simplify material identification. Macro-stage labels (slab, wall, roof, transitional) were manually assigned to each frame for training. Notably, the macro-stage prior (slab, wall, roof) only requires a coarse estimation of material distribution, as it operates on proportional material availability rather than precise object-level detection or counting. Future work may replace color-based segmentation with learned material recognition models while retaining the same material-distribution-based representation. Using OpenCV, the proportional HSV color composition of visible materials was computed at each time step. The seven micro-state ratios were then aggregated into three macro-stage features(Slab_ratio, Wall_ratio, Roof_ratio) forming a normalized macro-stage vector. Frames exhibiting high entropy or no dominant material ratio, corresponding to transitional or ambiguous contexts, were excluded during training to ensure stable macro-stage prior estimation.

Using valid frames from the three physical experiments, a probabilistic macro-stage prior P(macro|stock) was trained from the extracted macro stock vectors. For each macro stage (slab, wall, roof), a multivariate Gaussian distribution was fitted using supervised annotations. Cross-session generalization was evaluated using a leave-one-experiment-out (LOEO) protocol, training on two sessions and testing on the remaining one (Figure 7). Across all splits, the resulting confusion matrices exhibit a consistent and structured error pattern. Roof stages are classified with near-perfect accuracy, while limited confusion occurs only between the adjacent slab and wall stages. This behavior reflects construction logic rather than frame-level noise. As shown in the material ratio plot (Figure 8), the model produces stage predictions that evolve smoothly over time and differ from single-frame ratio thresholds by preserving uncertainty during transitions.

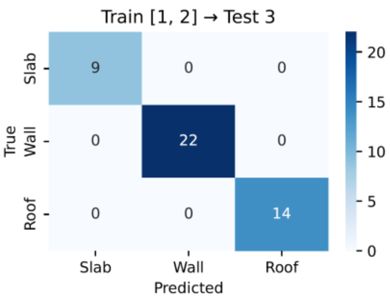

Figure 7. Confusion Matrix (Test Dataset 3)

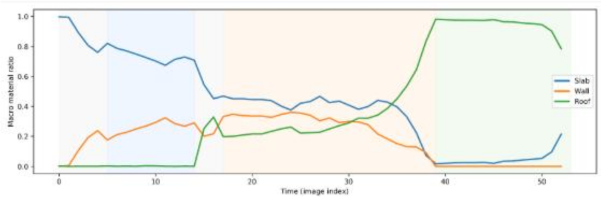

Figure 8. Material Ratio overlapped with Predicted Phase (Blue: Slab; Orange: Wall; Roof: Green)

Overall, material stock provides a stable and interpretable macro-stage prior that generalizes across sessions and explicitly represents ambiguity near stage boundaries. The model outputs a Gaussian macro-stage classifier, making it well suited for integration into higher-level, uncertainty-aware task inference.

2.4 Material-Aware Contextual Task Prediction

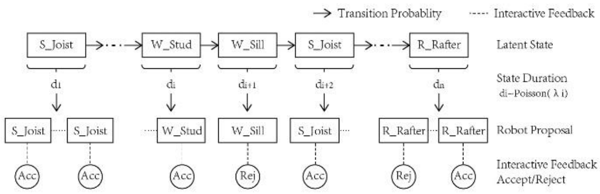

Figure 9. Online Interactive HSMM

Figure 9 illustrates the proposed online HSMM-based intention inference process with interactive feedback. The assembly process is modeled as a sequence of latent states, each corresponding to a component-level assembly action. State transitions are governed by the learned transition matrix A, while each state persists for a duration di drawn from a Poisson distribution parameterized by λ_i.

At each time step, the robot proposes the current latent state as a candidate assembly action. Human accept/reject feedback is incorporated online and directly modulates state duration rather than triggering an immediate state transition. Accepted proposals increment the duration of the current state, indicating continued execution, while rejected proposals partially consume duration with a penalty factor ρ, reflecting local correction without forcing premature state changes(Equation 1 & 2). This mechanism allows the model to remain within the same latent intent while adapting to human guidance.

$$\text{Accept: } d(s^t) \leftarrow d(s^t) + 1 \quad (1)$$

$$\text{Reject: } d(\hat{s}_t) + \rho d(s^t) \leftarrow d(s^t) + \rho \quad (2)$$

where $\rho<1$ penalizes rejected proposals.

Material context is incorporated through the pretrained Gaussian macro-stage classifier that estimates $P(m|x_t)$ from the current material stock vector. Each state is deterministically mapped to a macro stage m(s). State scores are computed by softly fusing temporal belief with the macro-stage prior:

$$score_t(s) = \pi_t(s) \cdot (P(m(s) \mid x_t) + \varepsilon)^{\alpha} \qquad (3)$$

$$with\ \alpha = 1.0\ and\ \varepsilon = 10^{-3}$$

This multiplicative bias promotes states consistent with the current construction phase while preserving uncertainty during transitions.

Finally, hard constraints remove structurally invalid successor states conditioned on the last accepted action. To ensure robust interaction, the system further suppresses repeatedly rejected states, penalizes short-term repetition, and introduces stochastic exploration when prolonged rejection is detected. Together, these mechanisms enable stable, intention-aware task prediction that adapts to both material conditions and human feedback throughout the assembly process.

3 Online HSMM-Based Material-Aware HRC Timber Assembly

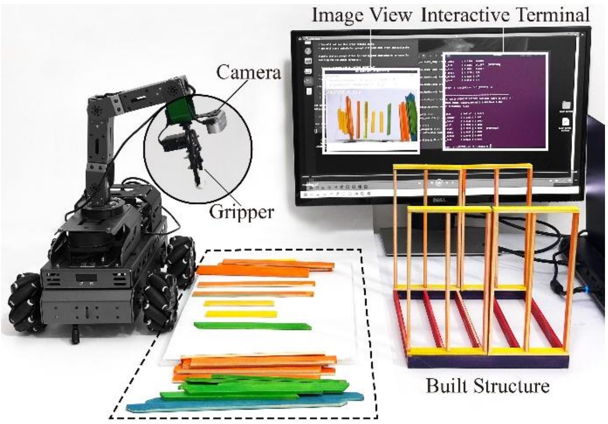

Figure 10. Picking and Handover Workplace

The human–robot collaboration (HRC) timber assembly system is implemented on a desktop robotic arm equipped with a parallel gripper and an RGB camera (Figure 10). The system is integrated through a ROS-based pipeline that supports perception, task inference, and interactive execution. After system initialization and camera calibration, visual features are extracted from the pick plane using color-based segmentation to estimate the spatial distribution of available timber components. These perceptual features are used to identify both the seven micro-level assembly states and their aggregation into three macro construction stages (slab, wall, roof), providing real-time material context for the subsequent task inference and handover process.

Unlike a conventional HSMM, observations in the proposed model are not emitted directly from hidden states. Instead, the HSMM generates candidate task hypotheses, which are evaluated through sparse human accept/reject feedback acting as interactive evidence for belief updates.

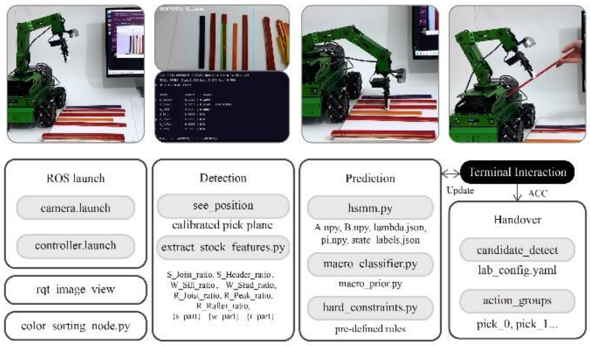

Figure 11. Online Material-Aware HRC Pipeline

Through this iterative interaction loop, intention inference is incrementally refined via human guidance rather than direct sensory emissions. As the assembly progresses, the robot continuously updates its belief by integrating material dynamics on site with interaction feedback, allowing emerging human construction patterns to be progressively captured during the collaborative process.

As shown in Figures 11 and 12, at each interaction step the robot first perceives the on-site material stock and proposes a candidate assembly state, which is displayed to the human builder. Human feedback is captured via a gesture recognition module implemented using MediaPipe. An open-hand gesture ("Five") indicates acceptance, upon which the robot retrieves the corresponding timber element from the pick area and hands it to the human. A closed-hand gesture ("Fist") indicates rejection, triggering an update of the state duration and a re-proposal of an alternative state.

Figure 12. Built Phases and User Interaction

4 Validation

4.1 Simulation and Comparison

1) To validate the proposed multi-model inference framework in a controlled experimental setting,

three configurations were evaluated on the same annotated timber assembly sequence: HSMM only: inference relies solely on temporal consistency and accept/reject feedback, without material or structural context.

2) HSMM + Macro: inference is augmented with a vision-derived macro-stage prior (slab, wall, roof) computed from material stock features and applied as a soft contextual bias over HSMM states.
3) HSMM + Macro + Hard: structural hard constraints derived from empirical construction logic are additionally enforced to eliminate physically invalid state transitions.

The HSMM-only baseline exhibits a strong imbalance between accept and reject events, with rejections far exceeding acceptances (Figure 13). Although the belief distribution becomes increasingly peaked, the most probable state is often misaligned with the true construction action, leading to repeated incorrect proposals and prolonged rejection loops—indicative of confident but incorrect inference driven solely by temporal consistency.

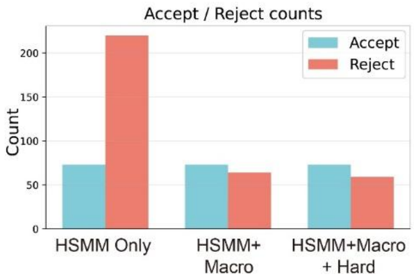

Figure 13. Accept/Reject Counts and Belief Entropy Results

Introducing a material-aware macro-stage prior (HSMM + Macro) substantially reduces rejection events while maintaining a similar number of acceptances. By biasing inference toward micro-states consistent with the current construction phase, the model avoids phase-inconsistent proposals (e.g., roof actions during slab or wall assembly), resulting in earlier belief convergence and more decisive actions. Adding hard structural constraints (HSMM + Macro + Hard) yields a smaller additional reduction in rejections, indicating that most incorrect proposals are already filtered by the macro prior, while hard constraints mainly provide incremental robustness by eliminating residual structurally invalid transitions.

Belief entropy analysis (Figure 14) supports these observations. The HSMM-only model maintains relatively high entropy over extended interaction periods, reflecting persistent uncertainty despite repeated proposals. In contrast, macro-augmented models exhibit faster entropy collapse, indicating earlier commitment to a coherent task interpretation. Accept/reject statistics further reveal a trade-off in interaction cost: HSMM-only distributes rejections over long horizons, while HSMM + Macro concentrates rejections into shorter intervals that expose high-level conflicts earlier.

The comparison results demonstrate that interaction feedback alone is insufficient for reliable intention inference. Material-aware macro-stage priors provide essential contextual grounding that allows human feedback to be interpreted meaningfully, enabling stable and efficient human–robot collaboration.

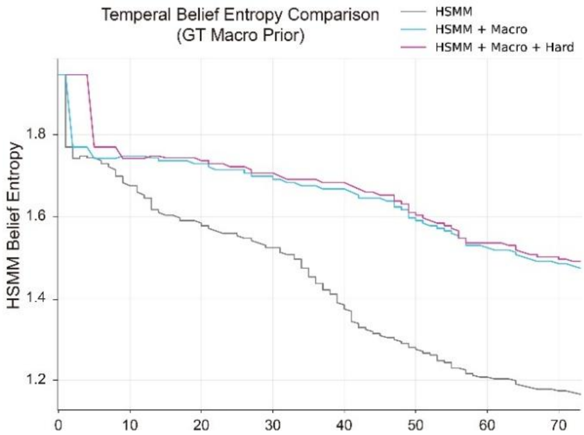

Figure 14. Temporal Belief Entropy Comparison

4.2 HRC Experiment Results

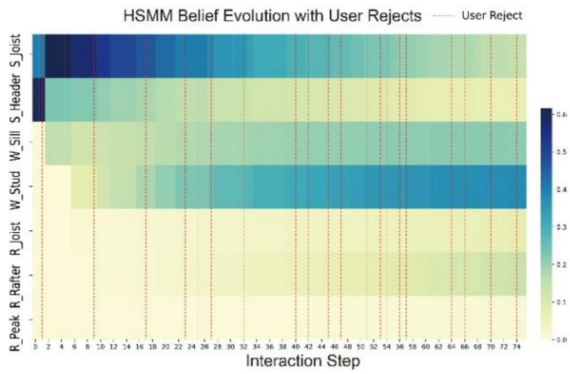

Figure 15. HSMM Belief Evolution of Experiment 1

The HSMM belief heatmap (Figure 15) shows that user rejection events systematically suppress incorrect state hypotheses while redistributing probability mass toward viable alternatives. Rather than inducing abrupt belief collapse, repeated rejections gradually reshape the belief distribution by consuming state duration and accelerating transitions, allowing recovery from initially dominant but incorrect hypotheses.

The confusion matrix (Figure 16) indicates that most prediction errors occur within the same macro construction phase (e.g., between W_Sill and W_Stud), while cross-phase confusions are rare. This confirms that the material-aware macro stage prior effectively constrains inference to the correct construction phase. During interaction, belief entropy(Figure 17) increases primarily in response to user rejections, reflecting deliberate reconsideration rather than overconfident convergence. Compared to the baseline, the HSMM + Macro + Hard variant converges more cautiously, supporting robust and safety-aware inference.

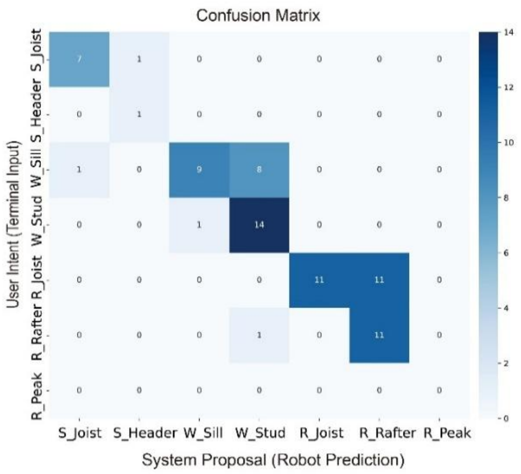

Figure 16. HRC Experiment 1 Confusion Matrix

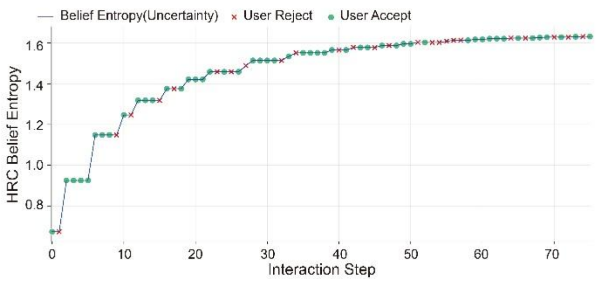

Figure 17. System Uncertainty and User Interaction of Experiment 1

Quantitatively, HSMM + Macro + Hard doubles the accept rate from 0.25 to 0.55 in simulation, closely matching the 0.58 observed in real HRC experiments(Table 1). The increased mean proposal duration (1.03 → 1.63 steps) indicates reduced rejection loops and improved interaction stability. These results demonstrate that sparse human feedback, when combined with material-aware priors, is sufficient for adaptive and interpretable task inference in construction contexts.

To evaluate scalability and generalizability, three timber assembly scenarios with increasing structural complexity were tested (Figure 18), including (1) a simple single-story frame, (2) a single-story structure with extended enclosure phases (longer wall-stage duration), and (3) a two-story configuration. As shown in Table 1, the proposed framework maintains consistently high performance across all scenarios, with stable accept rates and comparable interaction durations. This indicates that the material-aware macro-stage prior enables robust task inference under varying assembly sequences and material distributions. The second scenario further demonstrates that the model can adapt to extended phase durations that differ from the training data. The slightly lower accuracy in the two-story scenario is mainly associated with extended rejection sequences during macro-stage transitions, where overlapping material distributions between adjacent phases lead to temporarily ambiguous observations. Despite this, the system maintains stable interaction performance, as reflected by a comparable accept rate (51.4%) and mean interaction duration (1.96 steps), indicating that the framework remains robust under increased structural complexity.

Figure 18. Three Experiments of Three Typical Timber Structures (from left to right: Experiment 1, Experiment 2, Experiment 3)

Table 1. Mean Duration and Accept Rate Comparison

State	Mean Duration	Accept Rate
HSMM-Only(simulation)	1.03	0.25
HSMM+Macro+Hard (Simulation)	1.50	0.55
HSMM+Macro+Hard (HRC Experiment 1)	1.63	0.58
HSMM+Macro+Hard (HRC Experiment 2)	1.76	0.57
HSMM+Macro+Hard (HRC Experiment 3)	1.96	0.52

5 Conclusion

This paper addresses the challenge of enabling adaptive human–robot collaboration in timber construction, where assembly intent emerges from dynamic material conditions and experience-driven sequencing rather than predefined task models. To bridge this gap, we propose a material-aware contextual task inference framework that allows robots to interpret and anticipate human construction intentions under uncertainty. The work achieves three key objectives: (1) formalizing construction knowledge through a light-frame timber ontology with minimal hard constraints; (2) learning and inferring task intent by integrating temporal hidden semi-Markov models with material-aware macro-stage priors; and (3) demonstrating a human–robot collaborative workflow that adapts task proposals based on on-site material perception and human feedback. Together, these contributions advance intention-aware automation in construction by enabling adaptive, context-sensitive robotic assistance that captures tacit construction knowledge.

This work has several limitations that motivate future research. From a modeling perspective, HSMM inference becomes less reliable for rarely observed states with limited training data (e.g., unique components such as ridge peak elements), suggesting the need for minimal structural constraints or hierarchical state abstractions to stabilize prediction. From a data and scalability perspective, the current evaluation is limited to timber frame structure implemented in a laboratory-scale robotic setup. Future work will expand the dataset to include multiple timber structure typologies and construction scenarios, enabling further validation of scalability and supporting the learning of more generalizable material-driven assembly knowledge. In addition, the simplified color-based perception used in this study can be extended to incorporate richer material features such as timber type classification, color and texture descriptors, and material area estimation from top-view observations. With access to real construction data, these features could be extracted using object detection or segmentation models and used to compute material distribution statistics for training the macro-stage prior model.

References

[1] J. Adams, R. Bajcsy, J. Košecká, V. Kumar, M. Mintz, R. Mandelbaum, C.-C. Wang, Y. Yamamoto, and X. Yun. Cooperative material handling by human and robotic agents: Module development and system synthesis. *Expert Systems with Applications*, 11(2):89–97, 1996.

[2] M. Cramer, J. Cramer, K. Kellens, and E. Demeester. Towards robust intention estimation based on object affordance enabling natural human-robot collaboration in assembly tasks. *Procedia CIRP*, 78:255–260, 2018.

[3] R. Gomez, M. Sridharan, and H. Riley. What do you really want to do? Towards a theory of intentions for human-robot collaboration. *Annals of Mathematics and Artificial Intelligence*, 89(1):179–208, 2021.

[4] T. Herzog, J. Natterer, R. Schweitzer, M. Volz, and W. Winter. *Timber Construction Manual*. Walter de Gruyter, 2012.

[5] Z. Liu, Q. Liu, W. Xu, L. Wang, and Z. Ji. Adaptive real-time similar repetitive manual procedure prediction and robotic procedure generation for human-robot collaboration. *Advanced Engineering Informatics*, 58:102129, 2023.

[6] K. Nottensteiner, A. Sachtler, and A. Albu-Schäffer. Towards autonomous robotic assembly: Using combined visual and tactile sensing for adaptive task execution. *Journal of Intelligent & Robotic Systems*, 101(3):49, 2021.

[7] L. Peternel, N. Tsagarakis, and A. Ajoudani. Towards multi-modal intention interfaces for human-robot co-manipulation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2663–2669, 2016.

[8] P. Wang, H. Liu, L. Wang, and R. X. Gao. Deep learning-based human motion recognition for predictive context-aware human-robot collaboration. *CIRP Annals*, 67(1):17–20, 2018.

[9] W. Wang, R. Li, Y. Chen, Y. Sun, and Y. Jia. Predicting human intentions in human–robot hand-over tasks through multimodal learning. *IEEE Transactions on Automation Science and Engineering*, 19(3):2339–2353, 2022.

[10] H. Yu, V. R. Kamat, C. C. Menassa, W. McGee, Y. Guo, and H. Lee. Mutual physical state-aware object handover in full-contact collaborative human-robot construction work. *Automation in Construction*, 150:104829, 2023.

[11] R. Zhang, J. Lv, J. Li, J. Bao, P. Zheng, and T. Peng. A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations. *Journal of Manufacturing Systems*, 63:491–503, 2022.

[12] Y. Zhang, K. Ding, J. Hui, J. Lv, X. Zhou, and P. Zheng. Human-object integrated assembly intention recognition for context-aware human-robot collaborative assembly. *Advanced Engineering Informatics*, 54:101792, 2022.