A Multi-Camera Human Tracking Framework for Safe Human-Robot Collaboration in Construction-Site Robotic Arm Operations

Ren-Jie Li [1], Meng-Han Tsai [1], Liang-Ting Tsai [2], Yuxiang Chen [2] and Ci-Jyun Liang [3]

[1] National Taiwan University of Science and Technology, Taipei, Taiwan
[2]University of Alberta, Edmonton, Alberta, Canada
[3]Stony Brook University, Stony Brook, New York, USA
d11105003@mail.ntust.edu.tw, menghan@mail.ntust.edu.tw, liangtin@ualberta.ca, yuxiang.chen@ualberta.ca, cjliang@umich.edu

Abstract

The construction industry is characterized by a high degree of structural complexity, making it difficult to achieve a high level of modularization at the current stage. During operations involving robotic arms, human operators are often required to enter the robot's workspace at certain stages to perform fine adjustments, which may expose them to safety risks. Therefore, this study aims to develop a human-robot collaboration safety system framework that utilizes multi-view cameras to monitor, in real time, the positions of both the robotic arm and the operator. Object detection techniques are applied to streaming video for human tracking, while human pose estimation is used to analyze body posture. In addition, multi-view trajectory integration is employed to fuse information from different camera perspectives, enabling the system to determine whether the operator has entered a hazardous operating zone of the robotic arm. Ultimately, the proposed system provides real-time risk alerts while minimizing unnecessary interruptions to normal construction workflows, with emergency stop triggered when safety is compromised. This design reduces the likelihood of accidents caused by unintended entry into dangerous areas and enhances overall safety in human-robot collaborative operations.

Keywords -
Human-robot collaborative; Safety monitoring; Object detection; Human Pose Estimation

1 Introduction

The use of robotic arms has been widely adopted in many highly modular industries, such as automotive manufacturing, electronics assembly, and warehouse logistics. However, the introduction of robotic arms into the construction industry faces far greater challenges. In addition to the inherent complexity of construction projects, even components of the same type may vary in size, specifications, and construction requirements due to differences in design conditions, site constraints, and client demands. Moreover, to accommodate on-site conditions, design modifications are often made during the construction process. These high levels of uncertainty and environmental variability pose significant challenges to the integration of robotic arms into prefabrication and assembly workflows in construction.

In the early stages of design and planning, operators must repeatedly test processing workflows and related parameters to ensure production efficiency and product quality. At the same time, robotic arms used in construction can weigh several tons, and their end-effectors are often equipped with high-risk tools, such as nail guns and cutting machines, for timber cutting and assembly tasks. To prevent irreversible accidents caused by human error, this study proposes a safety framework for construction-site robotic arm operations. By integrating multi-camera human tracking and safety-zone assessment, the proposed framework not only improves the efficiency of human-robot collaboration but also provides enhanced safety supervision in robotic construction environments.

2 Literature Review

Human-robot collaboration (HRC) has gradually become a key trend in the development of smart manufacturing industries [1][2]. However, when operators and robotic arms work within the same workspace, potential safety hazards may arise, including collision, crushing, pinching, and puncture injuries. To prevent such safety incidents, a common protective strategy is to establish a protective field , typically implemented through the Safety-Rated Monitored Stop (SRMS) mode, in which a robot and its human partner are not allowed to operate simultaneously within the same workspace [3]. Under SRMS, personnel are restricted from entering the operational area while the equipment is active. This can be achieved using physical barriers or sensor-based solutions such as light detection devices. Once human presence is detected within the operational zones during robot operation, the system automatically stops the equipment. However, this approach inevitably reduces the effective shared workspace between the robot and the human operator, potentially constraining task flexibility and productivity. To overcome this limitation, an alternative collaboration strategy is the Speed and Separation Monitoring (SSM) approach, which allows humans and robots to move simultaneously within the shared workspace while maintaining a minimum protective distance at all times [3]. Such safety strategies rely on reliable perception of worker locations in real time. To capture worker locations in complex construction environments, multi-camera system plays an important role in a wide range of domains, including motion capture, surveillance, robotics [4]. Multi-camera people tracking (MCPT) is a challenging task, particularly in complex real-world environments, and is essential for the development of intelligent surveillance and monitoring applications [5]. This task also involves solving multiple computer vision challenges, such as object detection and multiple object tracking (MOT) [6]. However, most existing MCPT research has focused on general surveillance or crowd monitoring scenarios, and is often conducted in controlled or simulated environments. In contrast, HRC settings present additional real-world challenges such as frequent occlusions, limited camera viewpoints, real-time safety requirements, and the need to associate tracking results with robot operating zones. Therefore, a reliable multi-camera tracking framework tailored for HRC scenarios is required to accurately monitor human positions and safety-critical movements in real time. In this study, we develop a multi-camera human tracking and safety monitoring system that integrates object detection, MOT, and human-robot spatial interaction analysis to support SSM-based collaborative operations.

3 Method

This study develops a Multi-camera human tracking and safety-warning system framework designed for construction-site robotic arm operations (Figure 1). The system consists of three main modules: a tracking module, a projection module, and a safety-zone decision module. Two cameras are installed around the robotic arm workspace to capture on-site operation videos from multiple viewpoints. The tracking results are then projected onto a 2D ground-plane representation, enabling analysis of the relative spatial positions between workers and the robotic arm during operation. Based on the proposed three-stage safety-zone design, the system further determines the robot operating speed in a sequential manner, including full speed, reduced speed, and emergency stop.

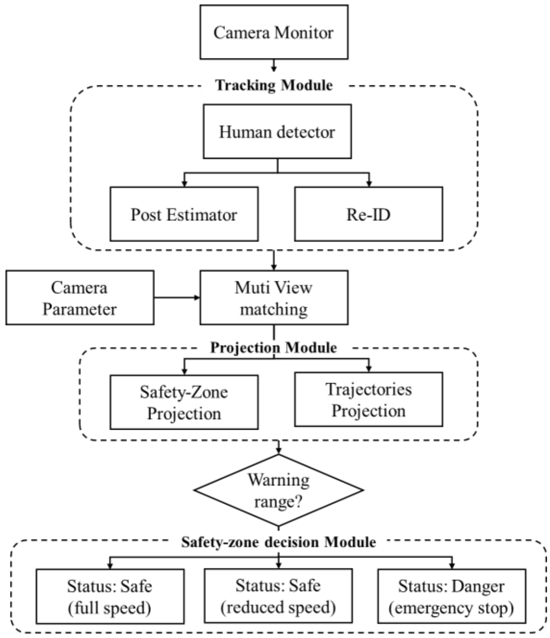

Figure 1. Multi-camera human tracking and safety-warning framework.

3.1 Tracking Module

The tracking module extracts identity-consistent worker trajectories from multi-view videos. First, each camera frame is processed using a YOLO-based detector to localize workers with bounding boxes. The model is fine-tuned using project-specific construction-site videos to improve robustness under real-world conditions. Next, a pose estimation model predicts 2D skeletal keypoints for each detection, providing structural cues that support cross-view association. Finally, person re-identification (ReID) features are extracted to match workers across cameras and time. By integrating appearance similarity, pose consistency, and temporal proximity, fragmented single-view tracks are fused into unified multi-camera trajectories.

3.2 Projection Module

In the projection module, fused multi-camera trajectories are mapped onto a 2D ground-plane representation of the workspace. Camera views are geometrically aligned with the site layout, and predefined safety-zone boundaries are projected onto the same ground-plane coordinate system. The system then computes worker-robot proximity by analyzing distances between workers and predefined hazardous regions within the robotic arm's working envelope. This projection supports real-time proximity monitoring and safety-aware decision-making under SSM-based human-robot collaboration scenarios.

3.3 Safety-Zone Decision Module

The safety-zone decision module evaluates operator proximity based on the projected 2D worker positions and predefined safety zones associated with the robotic arm's working envelope. Following the Speed and Separation Monitoring (SSM) principle, the workspace is divided into three staged zones reflecting different safety risk levels. The system continuously computes the relative distance between each worker and the hazardous operational region of the robotic arm. According to the zone in which the worker is located, the module determines the robotic arm operating mode sequentially:

1. Full speed operation when no workers are within the warning zones.
2. Speed reduction when a worker enters the intermediate zone.
3. Emergency stop when a worker enters the closest/high-risk zone.

This design provides an explicit safety supervision mechanism for real-time human-robot collaboration and reduces the likelihood of accidents caused by human error.

4 System prototype

To validate the feasibility of the proposed framework, a proof-of-concept prototype was developed with a focus on the tracking module. Since accurate human localization and identity-consistent trajectory generation provide the essential inputs for both ground-plane projection and safety-zone decision-making, tracking performance largely determines the effectiveness of the overall pipeline. To verify whether the estimated movement paths are consistent with real operator trajectories, a Unity-based simulation environment was constructed to replicate the robotic arm workspace (Figure 2)

The prototype implements multi-view tracking using three components: human detection, pose estimation, and person re-identification (ReID). Multi-view MP4 videos are decoded frame by frame, and a YOLO-based detector is applied to obtain worker bounding boxes. A top-down pose estimator predicts 2D keypoints for each detection, while a ResNet-50-based ReID model extracts appearance embeddings to maintain identity consistency across frames and camera views. The system then performs tracking integration by associating detections over time using fused cues from pose similarity, geometric constraints, bounding-box overlap (IoU), and ReID similarity.

Finally, for each confirmed track, the system projects the person's position onto a common 2D ground plane and compares the worker's location with the predefined safety-zone boundaries (Figure 3). When the operator is outside the robot work area, the robotic arm operates normally. Once the operator enters the fenced work zone, the robot is required to slow down. If the operator further enters the red danger zone, the robotic arm must stop immediately. This process is used to verify positional accuracy and assess whether the tracking result is consistent with the expected safety-zone relationship.

To quantitatively evaluate tracking accuracy, we measure the spatial error between predicted trajectories and ground-truth paths under different distance thresholds. The accuracy is defined as the percentage of frames where the localization error is within a given threshold. The results are summarized as follows: 0.1 m threshold: 26.3%, 0.3 m threshold: 80.4%, 0.5 m threshold: 88.9%, 1.0 m threshold: 95.3%. These results indicate that the proposed system achieves high localization reliability under practical safety margins (0.3-1.0 m), which are consistent with safety requirements in human-robot collaboration scenarios.

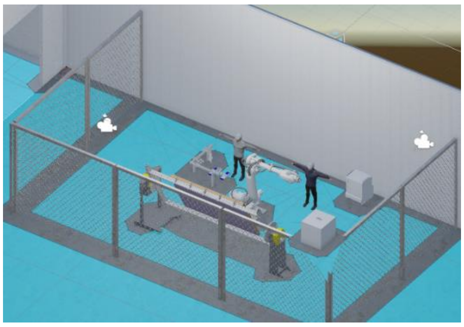

Figure 2. 3D on-site environment simulated in Unity.

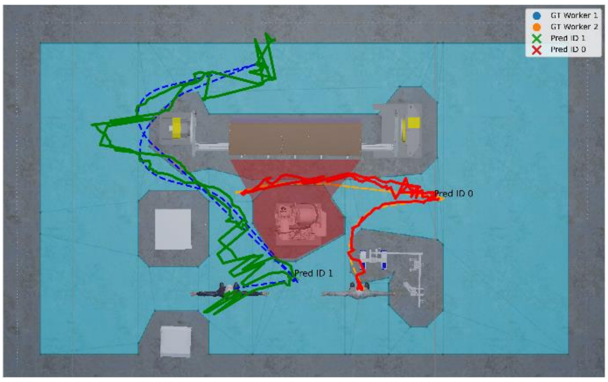

Figure 3. Human movement trajectories.

5 Future works

The proposed framework provides a safety supervision mechanism for human-robot collaboration in complex construction-site robotic arm workspaces. Compared with the traditional Safety-Rated Monitored Stop (SRMS) mode, which often reduces collaboration efficiency, this framework follows the Speed and Separation Monitoring (SSM) principle to gradually adjust the robotic arm operation state according to the actual safety risk. However, the current implementation adopts predefined safety zones for simplicity and real-time feasibility, which may not fully capture task-dependent hazards in dynamic construction environments. In practical scenarios, safety risks are influenced by multiple factors, including the robotic arm's reachability, tool path, and tool geometry. Therefore, future work will focus on integrating robotic arm motion planning and task-level information to dynamically model the working envelope of the robotic system. By incorporating task scripts and kinematic constraints, the system can generate adaptive safety zones that reflect real-time operational conditions. This enhancement will enable more accurate and context-aware safety assessment, improving the applicability of the proposed framework in real-world construction environments.

6 Conclusions

This study develops a Multi-camera human tracking and safety-warning system framework designed for application in robotic-arm operation environments within construction and factory settings. By deploying multiple video cameras around the robotic workspace, the system enables synchronized monitoring of operator activities and tracks human movement trajectories in relation to the robot's operating zone. The proposed framework is composed of three main modules: the tracking module, the projection module, and the safety-zone decision module. Through this pipeline, the system converts multi-camera visual information into a unified 2D spatial representation of worker positions, allowing the analysis of proximity and potential safety-risk interactions between operators and the robotic arm.

The proposed framework demonstrates the feasibility of using multi-camera human tracking as a safety-oriented perception layer for human-robot collaboration in construction-site robotic arm operations. By transforming tracking results into a spatial representation of worker-robot proximity, the system enhances situational awareness and supports collaborative operation strategies such as SSM, thereby contributing to safer interactions between workers and robotic systems. The system is designed to balance operational efficiency and safety by minimizing unnecessary interruptions while ensuring immediate emergency stop when hazardous conditions are detected, aligning with safety requirements in SSM-based human-robot collaboration.

Acknowledgements

The authors would like to thank Prof. Jacob J. Lin and Mr. Cheng-Yun Tsai (Department of Civil Engineering, National Taiwan University) for their support in prototype development, including coding assistance and constructive feedback. The authors also acknowledge RoBIM Technologies Inc. for sharing practical knowledge and providing support for this study.

References

[1] Occupational Safety and Health Administration, Ministry of Labor. Collaborative Robot Operation Safety Assessment Report (original in Mandarin). April 2019. Available at: https://www.osha.gov.tw/. Accessed on 2026/01/01.

[2] M. Valori, A. Scibilia, I. Fassi, J. Saenz, R. Behrens, S. Herbster, C. Bidard, E. Lucet, A. Magisson, L. Schaake, J. Bessler, G. B. Prange-Lasonder, M. Kühnrich, A. B. Lassen, and K. Nielsen. Validating Safety in Human-Robot Collaboration: Standards and New Perspectives. *Robotics*, 10(2):65, 2021. https://doi.org/10.3390/robotics10020065.

[3] W. Li, Y. Hu, Y. Zhou, and D. T. Pham. Safe Human-Robot Collaboration for Industrial Settings: A Survey. *Journal of Intelligent Manufacturing*, 35:2235-2261, 2024.

[4] S.-E. Lee, K. Nishino, and S. Nobuhara. *Spatiotemporal Multi-Camera Calibration Using Freely Moving People*. *IEEE Robotics and Automation Letters*, 10(5):4818-4825, May 2025. https://doi.org/10.1109/LRA.2025.3554102

[5] J. Kim, W. Shin, H. Park, and J. Baek. Addressing the Occlusion Problem in Multi-Camera People Tracking with Human Pose Estimation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 5463-5469, 2023.

[6] S. Hassan, G. Mujtaba, A. Rajput, and N. Fatima. Multi-Object Tracking: A Systematic Literature Review. *Multimedia Tools and Applications*, 83:43439-43492, 2024.