



UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

Safe and Efficient SSM Collaborative Strategy Integrating Human Head and Hands Tracking with a RGB-D Camera

This is the peer reviewed version of the following article:

Original

Safe and Efficient SSM Collaborative Strategy Integrating Human Head and Hands Tracking with a RGB-D Camera / Lettera, G., Scoccia, C., Callegari, M.. - ELETTRONICO. - (2025), pp. 7318-7323. (2025 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2025 Wien, Austria 5-8 October 2025) [10.1109/SMC58881.2025.11342686].

Availability:

This version is available at: 11566/362545 since: 2026-09-07T09:10:38Z

Publisher:

IEEE

Published

DOI:10.1109/SMC58881.2025.11342686

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.

This item was downloaded from IRIS Università Politecnica delle Marche (<https://iris.univpm.it>). When citing, please refer to the published version.

Publisher copyright:

IEEE - Postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. To access the final edited and published work see 10.1109/SMC58881.2025.11342686

(Article begins on next page)

Safe and Efficient SSM Collaborative Strategy Integrating Human Head and Hands Tracking with a RGB-D Camera

Gaetano Lettera, Cecilia Scoccia, and Massimo Callegari

Abstract—In the evolving Industry 5.0 landscape, human-robot collaboration (HRC) is crucial to improve safety and productivity in shared industrial workspaces. This paper presents a cost-effective solution to track the human skeleton and identify the body parts that are most exposed to collision risk during a HRC task, i.e. the head and the hands, using a RGB-D camera. The depth images are processed through artificial intelligence (AI) algorithms, ensuring an accurate estimation of human motion and proximity to the robot. Furthermore, a robot control module has been developed to modulate the robot's speed based on the risk assessment calculated in real time, implementing a collaborative Speed and Separation Monitoring (SSM) scenario. To improve the efficiency of the workcell, the research **proposes** a multiplication coefficient to compute the minimum protective separation distance suggested by the ISO 10218-2, which weights the actual risk based on the biomechanical limits of the human part involved, according to the maximum permissible forces provided by the standard. Experimental results in a real robotic work cell demonstrate that the proposed system can achieve safety performance comparable to traditional motion capture systems, such as the OptiTrack; meanwhile it does not require operators to wear additional equipment and offers the advantages of reduced costs and a simpler setup, which is more affordable and adaptable to a wider range of industrial environments.

I. INTRODUCTION

The rise of Industry 5.0 emphasizes the seamless integration of humans and robots in collaborative environments, promoting productivity while ensuring workers safety. In contrast to traditional automation, where robots operate in isolated spaces, HRC requires advanced safety measures to mitigate collision risks and optimize interaction efficiency. One of the main challenges in this area is the real-time tracking of human operators to ensure that robots adapt their behavior dynamically based on human movements. Ensuring safety in HRC is critical, particularly in shared workspaces where physical interactions between humans and robots happen frequently. Traditional motion capture systems, such as OptiTrack, have been widely used for accurate tracking of people. However, they have limitations in terms of cost,

infrastructure complexity and the need for operators to wear additional markers or suits.

Various approaches have been explored to enhance safety in HRC environments. Previous research [1], [2] introduced the concept of skeleton tracking for risk assessment but did not implement a fully functional system; Chen and Kia [3] demonstrated the effectiveness of OptiTrack for people tracking and security monitoring in collaborative workspaces. Although OptiTrack provides high accuracy, it requires a multi-camera configuration and dedicated infrastructure, limiting its applicability in dynamic or cost-sensitive industrial environments. RGB-D cameras have emerged as a cost-effective alternative for tracking people and objects in industrial environments [4], [5]: unlike marker-based systems, they provide depth perception without requiring wearable devices [6]. Recent advances in artificial intelligence-based vision systems enabled real-time pose estimation, making RGB-D solutions increasingly viable for HRC applications. However, challenges remain in terms of accuracy, occlusion management and compliance with industrial safety standards. Lettera et al. [7], [8], described an architecture for the safe execution of HRC industrial tasks and presented an experimental validation: the study effectively highlights the key challenges and provides application-oriented insights. However, while the work offers a solid foundation, it lacks a comparative analysis with alternative methodologies and a broader discussion on long-term implications. This paper proposes an alternative solution by introducing a new RGB-D-based tracking system that incorporates a SSM framework. By incorporating a biomechanical weighting factor to fine tune the calculation of the protective separation distance, the proposed method increases safety and optimizes the robot's operational efficiency. Experimental results show that this approach achieves comparable performance to OptiTrack while reducing configuration complexity and costs, making it a practical solution for widespread industrial adoption. By integrating AI algorithms, the system improves real-time risk assessment, allowing robots to dynamically adjust their speed and maintain a safe distance, as prescribed by the new ISO 10218 standard [9] and [10].

II. HUMAN TRACKING MODULE

This section describes the overall architecture of the proposed system, including the hardware and software components. Details about the 2D skeleton recognition software and the 3D head and hands identification are given to better understand the human-robot separation distance computation.

*This research was funded by PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR); Missione 4 "Istruzione e Ricerca" - Componente C2; Investimento 1.1, "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)" - D.D. n. 104 del 2 febbraio 2022, Bando PRIN 2022 - CUP I53D23001930006 - Cod. prog: 2022XXH9JZ_002 / ERC: PE8 - Project title: "DYNamic Assessment and Mitigation of the Impact of Collaborative Applications (DYNAMICA) <https://www.cobotics.it>.

G. Lettera, C. Scoccia and M. Callegari are with Dipartimento di Ingegneria Industriale e Scienze Matematiche (DIISM), Università Politecnica delle Marche, Ancona AN 60131, Italy g.lettera@staff.univpm.it, c.scoccia@staff.univpm.it, m.callegari@univpm.it

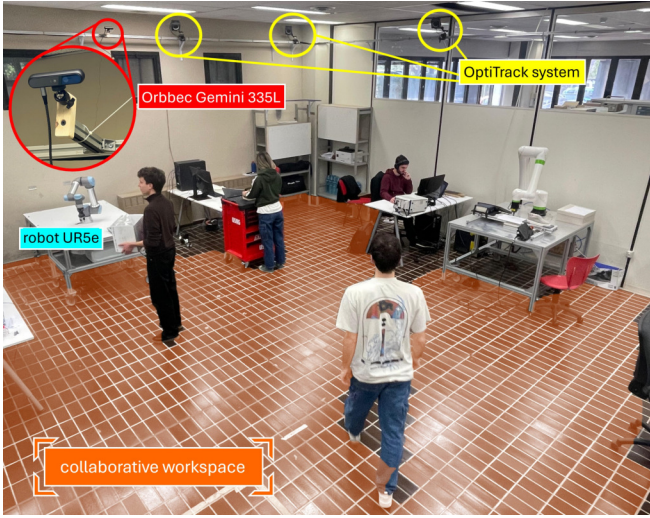


Fig. 1. The experimental setup composed of a UR5e robot, a depth camera (Orbbec Gemini 335L) and a motion capture system (OptiTrack).

A. Experimental setup

The experimental setup includes a Universal Robot UR5e and two perception systems (to be compared) that monitor the collaborative workspace: the OptiTrack system, installed along the perimeter of the entire room (as described in [7]), and the Orbbec Gemini 335L depth camera, installed in a fixed position with an orientation able to cover the field of view in front of the robot. The Orbbec Gemini 335L was chosen for its wide field of view ($87.5^\circ \times 58.0^\circ \times 101.0^\circ$), high-resolution imaging (1920×1080 , Full HD), and precise depth sensing (640×480). With a high frame rate (up to 60 FPS) and a depth range of up to 10 meters, it provides a cost-effective, flexible solution for real-time human tracking in large industrial settings.

When combining a depth camera into a robot architecture, the sensor has to be extrinsically calibrated to compute the relative pose between the adopted sensor and the robot base frame. This step can be performed by using a calibration target. In literature, this problem is solved by different calibration procedures, especially for object recognition applications. Their typical target is to recognize objects located as far as about 0.5 m from the camera frame. On the contrary, for the proposed experimental application, the robot and the operators work up to 10 m away from the camera. To obtain the desired accuracy, the depth camera extrinsic calibration procedure follows the one described in [11] more adapt for large workspace monitoring, exploiting a sphere tracking.

The developed software architecture is ROS2-based and runs on a desktop PC running Ubuntu 20.04 LTS with an Intel Core i9-14900KF CPU, 128GB of RAM and an nVidia RTX 4090 GPU with 24GB of memory.

B. Head and hands identification

To tackle the challenges of real-time human pose estimation, the proposed pipeline uses a Convolutional Neural Network (CNN) based on the YOLOv9 framework [12].

YOLOv9 is a state-of-the-art 2D object detector pre-trained on ImageNet [13] and fine-tuned on the MS-Coco dataset [14]. Known for its speed and accuracy, YOLOv9 can detect various objects, such as chairs, people, laptops, and buildings. In the standard YOLOv9 model, the input RGB image is processed and a 2D bounding box is created around each detected object with a corresponding label representing the object's class. In this research, the standard model has been retrained specifically to detect only the *person* class. Additionally, a 2D skeleton detection algorithm based on YOLOv9 is used to estimate the human skeleton model for each detected person, as shown in Fig. 2A which shows eleven operators into the collaborative workspace. The image is pre-processed to integrate depth and color data from the RGB-D camera, allowing the 2D detection outputs to be mapped into 3D space. A custom post-processing module refines these detections, integrating depth information to provide accurate 3D pose estimation and filters out low-confidence joint detections. For operator safety, the system focuses on detecting the head and hands, the body parts most exposed to collision risks in collaborative robot environments. A developed submodule tracks and visualizes the head and hands in the workcell digital mockup. Each operator is uniquely identified by an ID and a color, which facilitates the differentiation of the head and hands from multiple operators. The resulting detection is shown in Fig. 2B: each i -th operator is identified by three marker spheres, p_{ihead} , p_{irhand} and p_{ilhand} which represents respectively the 3D position of the operator head, his right hand and his left hand (in the figure only two operators have been defined for simplicity). Both the training and the prediction phases of this algorithm are computationally demanding and require a high-performance GPU (nVidia RTX 4090) for real-time processing and accurate results.

C. Human-Robot separation distance

A key aspect of implementing the SSM collaborative task is the real time computation of the separation distances between the robot and each operator. In the proposed pipeline, the positions of both the robot's end-effector, p_R , and the tracked human body parts, p_{ihead} , p_{irhand} and p_{ilhand} , are continuously monitored. To determine the relative distance, the algorithm computes in real time the Euclidean norm between p_R and the three operator markers, i.e. it computes the straight-line distance between two points, for example p_A and p_B , in 3D space. This is represented mathematically as $d = \|p_A - p_B\|$. This approach allows to identify not only the operator ID closest to the robot but also which part of the body is most exposed to risk due to proximity (head, right hand or left hand).

III. ROBOT SPEED SCALING ALGORITHM

This section explains the robot control module developed to adjust robot's speed based on the proximity and safety index. It describes how real-time vision data and a decision-making algorithm assess risk. The collaborative task follows the ISO 10218-2 standard, proposing a multiplication factor

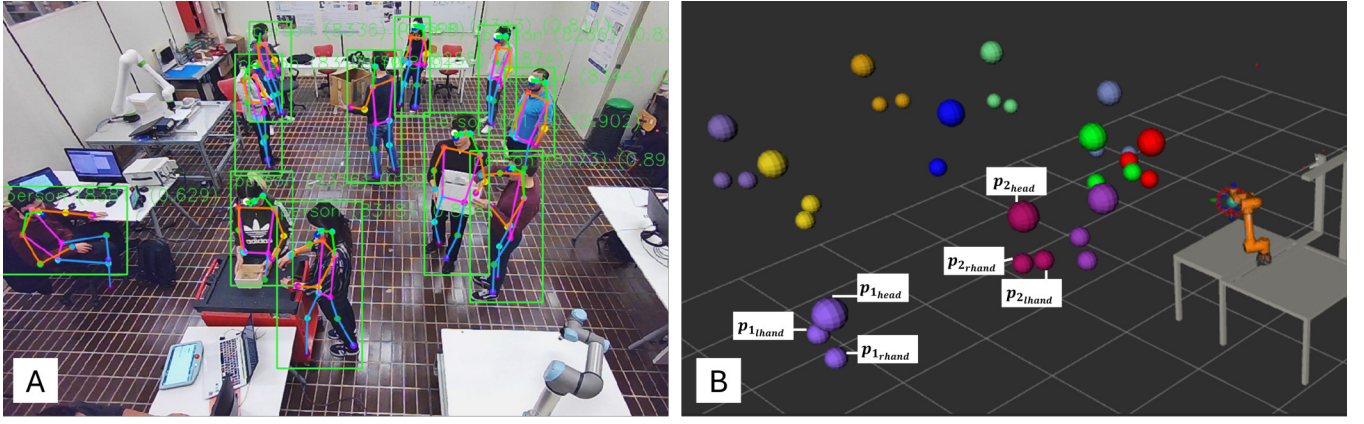


Fig. 2. Human tracking module: from 2D skeleton tracking (A) to 3D head and hands identification (B).

that adjusts the protective distance between humans and robots based on the biomechanical limits of the body part at risk.

A. Estimation of the human operator and robot velocities

A key part of HRC module is the intention estimation (IE), which predicts human movements. A linear Kalman filter (LKF) is used to estimate the velocity of the key human body points (p_{ihead} , p_{irhand} and p_{ilhand}) by combining depth camera data with a motion model. The filter predicts and updates the future position and velocity of each marker in real-time, enhancing accuracy. These estimates are used to estimate the safety index of the collaborative task:

$$x_{k+1} = \begin{bmatrix} I_3 & \Delta t I_3 \\ O_3 & I_3 \end{bmatrix} x_k + w_k, \quad (1)$$

$$y_k = \begin{bmatrix} I_3 & O_3 \end{bmatrix} x_k + n_k \quad (2)$$

where Δt is the sampling time, I_3 and O_3 are the identity and zero matrices of size 3×3 , respectively; w and n are the process and measurement noises with covariance matrices W and N , respectively. Finally, x is the state vector of the system, i.e., the position and the velocity of the human marker $x = [p_i^T \ \dot{p}_i^T]^T$, and the measured output y is a vector containing the coordinates of the point p_i described in Section II-B. The LKF equations used in this work are standard and are not included here for the sake of brevity.

On the other hand, the robot end-effector position, p_r , is computed in real time by retrieving the transform between the robot base frame, T_{base} , and the end-effector frame, T_{ee} , which are updated in real time by the robot controller and drivers. The robot end effector velocity, v_r , is computed by evaluating the change in the end effector position over time. More in detail, the change in position is calculated between current and previous positions of p_r , while the time difference, dt , is calculated between the current and previous timestamps of the transform.

Fig. 3A shows three human operators moving in the collaborative scenario while the robot is running a pick and a

place operation: the first operator carries a box and walks in the direction of the robot, the second operator moves away from the robot, and the third operator is almost stationary performing an inspection operation close to the robot. The results of the developed LKF software are shown in Fig. 3B. The velocities of the three points p_{ihead} , p_{irhand} and p_{ilhand} of i -th operator are represented by their own components (arrows following the RGB convention) and by the resulting velocities v_{ihead} , v_{irhand} and v_{ilhand} (yellow arrows). Their length is scaled according to their module: the velocities of the walking operators are significantly higher than the velocities of the third operator who is almost stationary. Fig. 3B also shows the robot end-effector positions (purple sphere) and velocities (visualized arrows) during its motion.

The final step of the proposed IE module involves calculating the relative speeds between the humans and the robot. These speeds indicate how quickly the human and robot are approaching or moving away from each other, a critical factor in determining the protective distance S , as outlined by ISO 10218-2 [10]. The human (to robot) directed speeds, v_{ihead}^d , v_{ilhand}^d and v_{irhand}^d , represent how fast the markers of the i -th human operator are approaching or receding from the robot. They are computed by projecting the velocity vector onto the direction from the specific marker position to the robot's position. For example, for the i -th human head, v_{ihead}^d is the dot product of the velocity vector v_{ihead} and the direction vector $d_{ihead \rightarrow r}$, which quantifies the extent to which the human head's movement aligns with the direction toward the robot:

$$d_{ihead \rightarrow r} = \frac{p_r - p_{ihead}}{\|p_r - p_{ihead}\|}, \quad v_{ihead}^d = v_{ihead} \cdot d_{ihead \rightarrow r} \quad (3)$$

A positive value for v_{ihead}^d indicates that the head is approaching the robot, whereas a negative value means they are moving away. The robot (to human) directed speed is defined accordingly.

These directed speeds are essential for determining the protective distance S between each human marker and the

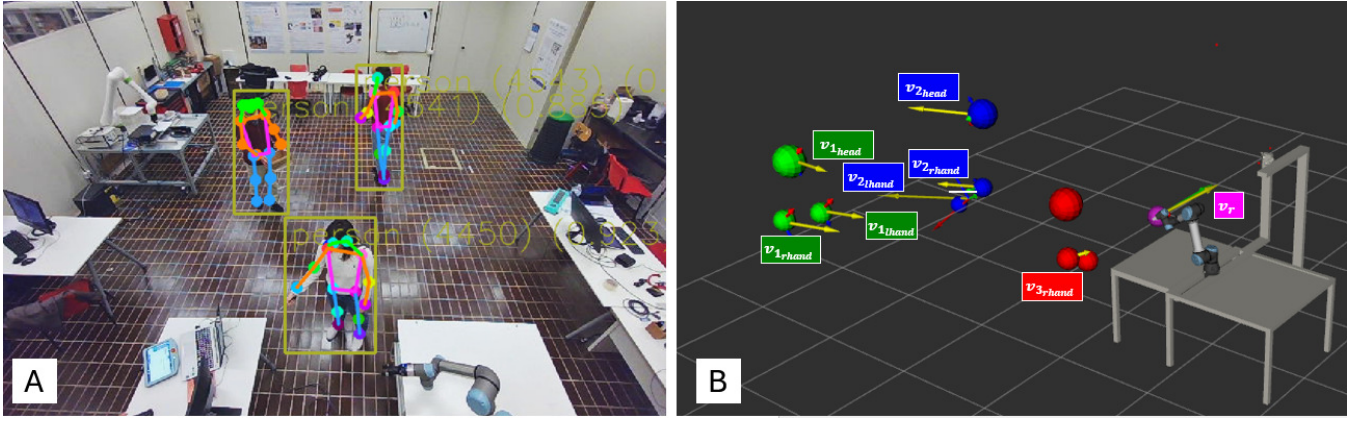


Fig. 3. Pipeline for human operator and robot velocities estimation: (A) identification of the three operators moving in the collaborative workspace and (B) estimation of the velocities of their corresponding heads and hands.

robot. This distance depends on how quickly both the human and the robot are closing the gap between each other. Specifically:

- If v_{thead}^d is large and positive, it indicates that the human head is moving quickly towards the robot. This could reduce the time available for the robot to react or stop to avoid a potential collision.
- Similarly, if v_r^d is large and positive, it means that the robot is rapidly moving towards the human, shortening the time available to avoid the contact.

By factoring in these velocities, S can be adjusted to ensure that the robot has enough time to respond and maintain a safe distance based on the relative speeds of both entities.

B. Weighted protective separation distance

The protective separation distance S in Annex L of the current ISO 10218-2 standard [10] does not account for the specific body part at risk, treating all parts the same despite their different biomechanical importance. **The proposed system reduces the robot's speed when high-risk body parts are involved, and allows higher speeds when lower-risk parts are detected, improving both safety and productivity. The research introduces a multiplicative factor, γ , which associates different weights to different body areas, according to Tab. M.2 of Annex M in [10].** The minimum protective separation distance S is therefore redefined as:

$$S^\gamma = \gamma[(v_i T_r + v_i T_s) + (v_r T_r)] + (B) + (C + Z_s + Z_r), \quad (4)$$

γ is calculated proportionally to the maximum force limits allowed by the ISO 10218-2 [10]. Specifically, from the tabulated values, the human head has the lowest limit at $f_{min} = 65N$, while the highest value is associated with the thigh, at $f_{max} = 220N$. The proposed approach assigns $\gamma(f_{min}) = 1$ if the involved body part is the head and $\gamma(f_{max}) = 0.5$ if the involved body part is the thigh. For other body parts, γ values ranging between 0.5 and 1. The value of γ is thus calculated based on the maximum force

value, f , listed in the table and it is calculated using the following interpolation equation:

$$\gamma(f) = \gamma(f_{min}) + (\gamma(f_{max}) - \gamma(f_{min})) \frac{f - f_{min}}{f_{max} - f_{min}}. \quad (5)$$

For the hands, with a maximum force of 140N, $\gamma = 0.742$.

C. Robot speed modulation

The last step of the robot control module is to compare d_i and S_i , associated with each tracked i -th marker, as suggested by the actual regulations. The comparison computes the robot speed override (or scaling factor), k , which adheres to the SSM principle. The speed override is defined as follows:

$$k = \begin{cases} 1.0 & \text{if } d_i \geq \Delta S_{max}^\gamma \\ 0.0 & \text{if } d_i < S_{max}^\gamma \text{ or if } d_i < d_{min} \\ \frac{d_i - S_{max}^\gamma}{\Delta \times S_{max}^\gamma - S_{max}^\gamma} & \text{otherwise} \end{cases} \quad (6)$$

where S_{max}^γ is the maximum value in the S_i^γ array, Δ is a constant scaling factor that determines the range of distances for which the speed override is computed (it is chosen as 2) and d_{min} is a minimum distance threshold that has been introduced to stop the robot if the operator is extremely close to the robot itself (it is chosen as 0.30 m). The first condition represents a totally safe condition for the human operators, thus the robot can move to its maximum speed while the second condition satisfies the regulation ISO 10218-2. For intermediate distances, the proposed strategy uses a linear interpolation to gradually shape the scaling factor between 0.0 and 1.0 based on the distance.

IV. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed Orbbec-based system, comparing its safety and efficiency performance with the OptiTrack-based system [7].

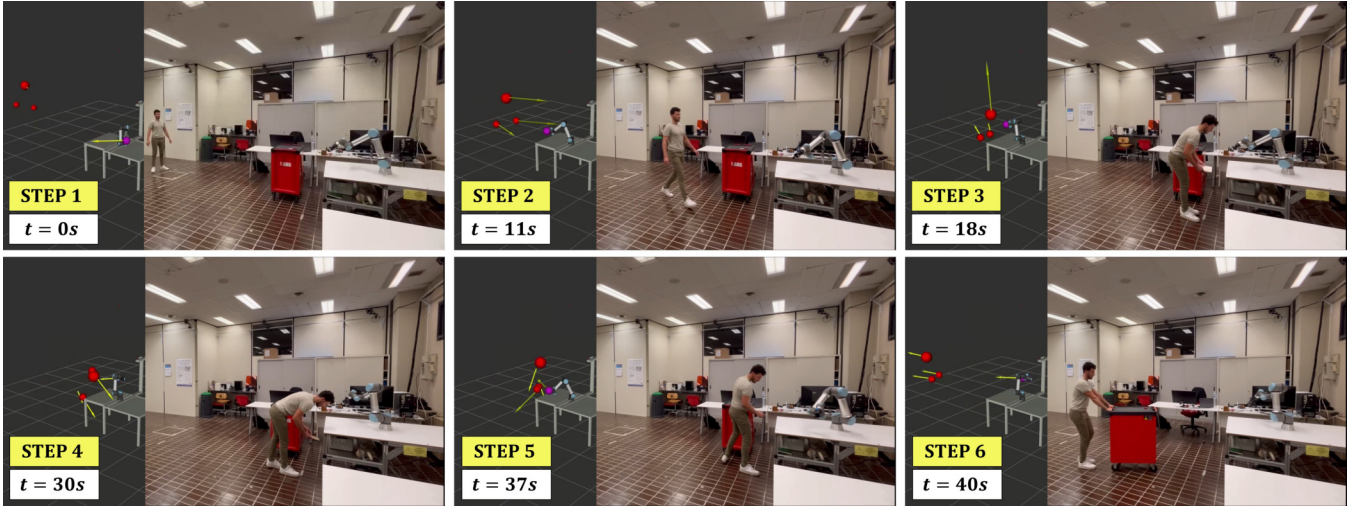


Fig. 4. Snapshots of the task experiment.

1) *Task description:* The scene simulates a robot executing a pre-planned path at a given nominal speed and a human operator entering the collaborative workspace to perform some inspection and maintenance activities. Fig. 4 shows the main steps of the task execution, while the accompanying video is available at this link¹. At the beginning (STEP 1), the operator is 6 meters away from the robot and the robot moves at the maximum predefined speed. At $t = 11$ s the operator walks towards the robot station (STEP 2), then slows down and stops in front of the desk. During STEP 3, the operator bends down to pick up the notepad under the robot table, then stands upright (at $t = 18$ s) to start an inspection task. At $t = 30$ s the operator leans forward to place the notepad under the robot table, then the robot stops (STEP 4) and the robot continues its trajectory only when the operator steps back. At STEP 5 the operator takes a screwdriver from the trolley to perform a assembly task, causing the robot to slow down and stop. Finally, the operator moves away from the robot at about $t = 40$ s (STEP 6).

2) *Task results:* Results are shown in Fig. 5. The top graph compares the speeds $v_{1_{head}}$ (red line, operator head), $v_{1_{rhand}}$ (orange line, operator right hand), $v_{1_{lhand}}$ (green line, operator left hand) and v_r (violet line, robot end effector). The bottom graph represents the computed distances which are processed to compute the robot speed scaling factor, k . Note that, for graphical clarity, the graph shows S_{max}^{iso} (dark blue line) and S_{max}^{γ} (light blue line). S_{max}^{γ} is compared with the respective distances d_{head} , d_{rhand} and d_{lhand} : the graph shows d (purple line) which is the distance of the marker corresponding to the one which makes the maximum value of S^{γ} . The minimum allowed distance d_{min} is represented by the red line.

More in detail, during STEP 1, the operator is stationary ($v_{1_{head}} \approx v_{1_{lhand}} \approx v_{1_{rhand}} \approx 0$ m/s), so $S_{max}^{iso} \approx S_{max}^{\gamma}$ and

they only depend on v_r : $d \geq \Delta S_{max}^{\gamma}$, thus $k = 1$.

STEP 2 clarifies the advantage of using S_{max}^{γ} in the place of S_{max}^{iso} in terms of robot downtime: the safety system is more permissive when the part of the human body subject to risk is not the head but the hands. At about $t = 11$ s the operator walks towards the robot, thus the operator speeds increase. Since the direction of motion of the operator is opposite to the direction of motion of the robot, S_{max}^{iso} and S_{max}^{γ} increase. Note that $S_{max}^{\gamma} < S_{max}^{iso}$, since the movement of the hands is faster than the head. During the walk, d decreases, until it becomes very similar to S_{max}^{γ} : the control system slows down v_r , until $k = 0$ when $S_{max}^{\gamma} > d$. The control system would have slowed down the robot much earlier if it had considered S_{max}^{iso} , while with the proposed system a gain in productivity was obtained.

STEP 3 highlights that the proposed system is conservative for hazards related to the head. At the beginning of the phase, the operator is leaning his head towards the robot, thus $S_{max}^{\gamma} \approx S_{max}^{iso}$: the robot slows down when it moves in front of him to prevent possible risky situations. When the operator retracts the head to an upright position, $v_{1_{head}}$ increases (red line, $t \approx 19$ s): the motion direction is opposite with respect to the robot motion, thus S_{max}^{γ} decreases and the robot resumes its normal speed v_r .

Similar to STEP 3, also STEP 4 demonstrates that the control system is more conservative when the operator approaches the robot with his head. STEP 5, instead, demonstrates that if the vision system identifies the hands as the parts of the human body most at risk of collision (maintenance activities), the safety system is more permissive thus $S_{max}^{\gamma} < S_{max}^{iso}$.

Finally, during STEP 6, at $t \approx 41$ s, the operator moves away from the robot station. He moves in the same direction than the robot, thus S_{max}^{γ} decreases and the collaborative scenario is strongly safe, thus $k = 1$.

¹Accompanying video: https://www.dropbox.com/scl/fi/7a9m4qtt5z9gujxga99s9/SMC2025_video_final_1.mp4?rlkey=ckffmt3wxqajxd49rlvsnl26f&dl=0

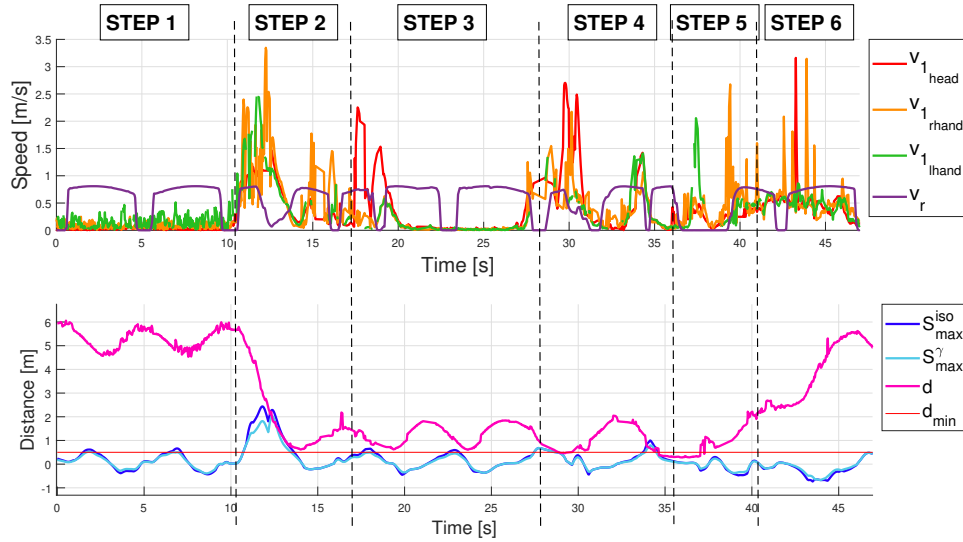


Fig. 5. Results of the task experiment.

V. CONCLUSIONS

This research presents a cost-effective solution to improve HRC in industrial settings by using RGB-D cameras and AI algorithms to track human body parts (head, hands) and assess collision risks in real-time. The robot's speed is dynamically adjusted based on these risk assessments, following ISO 10218-2 standards. By factoring in biomechanical limits of body parts, the system offers a more context-sensitive approach to safety. Compared to traditional systems like OptiTrack, it is simpler, more adaptable, and eliminates the need for additional equipment, making it suitable for cost-sensitive environments.

The performance of the proposed RGB-D camera-based system has been qualitatively compared with the results obtained by using the OptiTrack motion capture system described in the previous research work [7]. The tracking accuracy and system performance of both setups were found to be comparable. The proposed system, based on a single RGB-D camera, is more flexible and cost-effective but may introduce occlusions. **Current developments rely on a two-camera depth setup to resolve occlusions, but its evaluation is still ongoing and is not included in this paper.** Future work will include additional cameras to mitigate this limitation, incorporating the tracking of more body parts (torso, legs) in the risk assessment for a more comprehensive safety evaluation. Hands and head positions will be compared to assess tracking accuracy, while the safety index and corresponding robot speed adjustments will be analyzed to validate the proposed system through the OptriTrack.

REFERENCES

- [1] H. Huang, H. Hu, F. Xu, Z. Zhang, and Y. Tao, "Skeleton-based automatic assessment and prediction of intrusion risk in construction hazardous areas," *Safety science*, vol. 164, p. 106150, 2023.
- [2] H. Chen, X. Luo, Z. Zheng, and J. Ke, "A proactive workers' safety risk evaluation framework based on position and posture data fusion," *Automation in Construction*, vol. 98, pp. 275–288, 2019.
- [3] C. Chen and S. S. Kia, "Optitrack-aided supervised learning for neural network-based ultra-wideband ranging bias correction," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8484–8492, 2024.
- [4] T. Tashtoush, L. Garcia, G. Landa, F. Amor, A. N. Laborde, D. E. Oliva, and F. G. E. Safar, "Human-robot interaction and collaboration (hri-c) utilizing top-view rgb-d camera system," 2021.
- [5] O. Mazhar, B. Navarro, S. Ramdani, R. Passama, and A. Cherubini, "A real-time human-robot interaction framework with robust background invariant hand gesture detection," *Robotics and Computer-Integrated Manufacturing*, vol. 60, pp. 34–48, 2019.
- [6] A. P. Dani, I. Salehi, G. Rotithor, D. Trombetta, and H. Ravichandrar, "Human-in-the-loop robot control for human-robot collaboration: Human intention estimation and safe trajectory tracking control for collaborative tasks," *IEEE Control Systems Magazine*, vol. 40, no. 6, pp. 29–56, 2020.
- [7] G. Lettera, D. Costa, and M. Callegari, "A hybrid architecture for safe human–robot industrial tasks," *Applied Sciences*, vol. 15, no. 3, p. 1158, Jan. 2025. [Online]. Available: <http://dx.doi.org/10.3390/app15031158>
- [8] G. Lettera, A. Bajrami, D. Costa, and M. Callegari, *Dynamic Safety Evaluation and Risk Mitigation Strategies for Collaborative Kitting*. In *New Trends in Mechanism and Machine Science*; Springer Nature Switzerland, 2024, p. 286–293. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-67295-8_32
- [9] International Organization for Standardization, "Iso 10218-1:2025 - robotics — safety requirements — part 1: Industrial robots," 2025. [Online]. Available: <https://www.iso.org/standard/73933.html>
- [10] —, "Iso 10218-2:2025 - robotics — safety requirements — part 2: Industrial robot applications and robot cells," 2025. [Online]. Available: <https://www.iso.org/standard/73934.html>
- [11] M. Costanzo, G. De Maria, G. Lettera, C. Natale, and D. Perrone, "A multimodal perception system for detection of human operators in robotic work cells," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 692–699.
- [12] C.-Y. Wang and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.13616>
- [13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009, pp. 248–255. [Online]. Available: <http://www.image-net.org/papers/imagenet.cvr09.bib>
- [14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, pp. 740–755.