A human-robot collision avoidance method using a single camera

Ziyang Xie, Lu Lu, Hanwen Wang, Li Li, Xu Xu
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University

Human-robot collaboration is a flourishing work configuration in modern plants. Yet, the potentially hazardous collision between human workers and collaborative robots raises safety concerns. In this study, we proposed a collision avoidance method in which a single camera and a computer-vision algorithm were deployed to sense the location of human workers. Two collision avoidance schemes were further developed to determine the timing for robot to retract its arm. Specifically, the static scheme continuously monitors whether a worker is in a hazard zone, while the dynamic scheme predicts worker's position after a short time, and monitors whether the predicted worker's position is in a hazard zone. Preliminary validation showed that our proposed method can effectively enable a collaborative robot to retract its arms when a worker is approaching.

INTRODUCTION

Human-robot collaboration is a blooming work configuration in which human workers and collaborative robots (co-robots) work together in a shared workspace. This configuration takes advantage of the co-robot's endurance and the human's ability to react to unpredicted and less structured environments (Murashov et al., 2016). For instance, human-robot collaboration is commonly observed in the automobile industry where co-robots handle heavy parts while workers need to guide the co-robots' behaviors to ensure the co-robots work properly (Michalos et al., 2014) (Figure 1). It is expected that the market of co-robot will proliferate at a compound annual growth rate of 61% (Sherwani et al., 2020).

Figure 1. Two human-robot collaboration tasks performed in our Automation Lab. Left: A co-robot holds a deformable assembly base and the student mounts parts using a drill. Right: the student uses a pneumatic device to carefully level an assembly base, and the co-robot mounts small parts on this base. Given the nature of collaborative tasks, physically isolating the human workers from the co-robots is not applicable.

Since a co-robot works with human workers, it is no longer an option to isolate the workers from the co-robot for avoiding potential collisions. Thus, the potentially hazardous collision raises safety concerns in human-robot collaboration. To date, multiple engineering design features, such as limited end effector speed, torque sensors (Schäffer et al., 2008), and rounded exterior (Matthias et al., 2011), have been implemented in the industrial co-robot design to improve working safety (Matthias et al., 2011). Yet, the National Institute of Occupational Safety and Health (NIOSH) (NIOSH, 2017) reported that the overall risk of human-robot collaboration is still concerning. Particularly, as co-robots were just introduced to the industry in recent years, human workers have limited experience of working with co-robots. This results in a lack of understanding of the potentially hazardous co-robot behavior. For example, workers may not fully understand the moving trajectory of a co-robot's end effector and may collide with it.

One possible way to avoid the collision is to inform a co-robot of workers' positions and postures and enable the co-robot to actively respond to workers' movement (Michalos et al., 2014). The key to this method is to track workers' motion. To date, there are various methods for human motion tracking, and one of the most robust and accurate measurement system is laboratory-grade optical motion tracking system, which includes multiple cameras with markers attached to a human body (Furtado et al., 2019). However, applying such a motion tracking system in a plant for worker's body motion tracking can be challenging because of the bulky size and the high cost of the motion tracking system as well as the expertise required for running the system. An inertial measurement units (IMU)-based motion tracking system is also able to track human body motion and has been applied in different studies (Prayudi & Kim, 2012; Xie et al., 2021). Yet, for a precise body motion tracking, an IMU sensor needs to be attached on each body segment, which could be time consuming and may affect worker's natural body motion.

Alternative methods of human motion tracking for collision avoidance have been also reported in recent years. For instance, capacitive proximity sensors can detect the change of electric field when a worker is approaching. Thus, these sensors can be attached to a robot arm to detect nearby workers and help the robot arm respond to potential collisions (Schlegl et al., 2013). However, capacitive proximity sensors show limitations in the object detection range and the sensitivity to electrical noise. Multiple depth cameras have also been applied to collision avoidance (Mohammed et al., 2017; Schmidt & Wang, 2013). Depth cameras have good precision in nearby object detection. However, they are less robust for recognizing a worker from the background, which is essential in human-robot interaction. In addition, one needs at least two depth cameras to reconstruct 3D augmented environment due to the limited view angle of the depth sensor, which increases the difficulty of the field application.

In recent years, state-of-the-art open-source computer-vision algorithms, such as OpenPose (Cao et al., 2019), Detectron2 (Girshick, 2019) and VideoPose3D (Pavllo et al., 2019), have been developed to extract 2D and 3D human body posture from an image or a video clip captured by a regular RGB camera. These algorithms can recognize a human from a complicated background, which may solve the potential problems associated with depth sensors. Furthermore, the 3D pose reconstruction accuracy of Videopose3D is comparable with depth sensors (Pavllo et al., 2019; Plantard et al., 2015). In addition, compared with a motion tracking system, a regular camera is much less costly and can be easily deployed around a co-robot in the field. Therefore, a regular RGB camera may serve as another alternative for achieving body motion tracking in human-robot collision avoidance.

In this study, we focused on integrating a single RGB camera with a co-robot and developed proof-of-concept collision avoidance schemes during human-robot collaboration using images collected by the camera. The rest of the paper is organized as follows. The method section described the apparatus, data flow, and validation method of the proposed collision avoidance schemes. The discussion section briefly discussed the validation outcomes, current limitation and future works.

METHOD

Apparatus

A co-robot (Sawyer, Rethink Robotics) with 7 degrees of freedom was adopted in this study (Figure 2). The Sawyer robot was connected to a workstation with an NVIDIA RTX 2080Ti GPU to support the deployment of computer-vision algorithms. All the programming was done using Python (ver 3.6) in Linux Ubuntu 16.04. The robot control was realized by using Robot Operating systems (ROS Kinetic) framework. A 100° wide-angle webcam (Spedal, MF920P) was placed by the Sawyer robot and connected to the workstation. The placement of the camera ensured that workers interacting with the co-robot can be fully covered.

Real-time worker pose reconstruction

The real-time human pose reconstruction was realized by processing the image frames captured by the camera successively. Each frame was input to Detectron2 (Girshick, 2019), an open-source computer-vision package that is able to identify the 2D key joints of the human body in an image. The identified body joint locations in the image were then input to VideoPose3D, another open-source package that can reconstruct 3D human pose in camera space based on the 2D key body joint locations (Figure 2). With the workstation adopted in this study, the time of analyzing one video frame is 0.17 seconds, which is equivalent to a sampling rate of 5.8 Hz.

Coordinates calibration

Note that the 3D coordinates of key body joints output from Videopose3D is in camera space. In order for the co-robot to interpret these coordinate values, these values need to be converted to a coordinate system that the robot can understand. Thus, a calibration process is needed. For the convenience of algorithm development, the origin of this new coordinate system was set to the worker's head during the calibration as shown in Figure 3, and this new coordinate system was referred as work space hereafter.

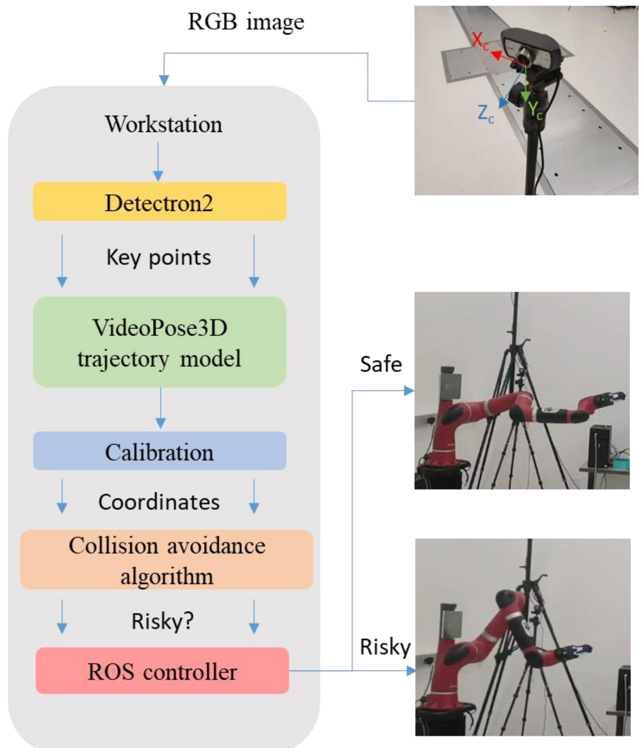

Figure 2. The workflow of the proposed collision avoidance method. Directions of each axis in camera space are marked on the webcam.

In work space, x- and z-axis form the the horizontal plane, and y axis is the axis perpendicular to the horizontal plane. During the calibration, a worker stands in front of the collaborative robot at an arbitrary location, and the worker's head coordinate derived from VideoPose3D is $\mathbf{P_{0c}} = (X_0, Y_0, Z_0)_c$ in camera space, which corresponds to $\mathbf{P_{0w}} = (0,0,0)_w$ in work space. Initially, the worker stands straight (Figure 3c), and the coordinate of middle point of the worker's two ankles is recorded as $\left(X_y, Y_y, Z_y\right)_c$. The worker's height is measured as L_y. The worker then moves half a meter ($L_x = 0.5\ m$) in x-axis (left), and the recorded head's coordinate in camera space is $(X_x, Y_x, Z_x)_c$. Next, the worker moves back to the original location and then moves half a meter ($L_z = 0.5\ m$) in z-axis (backward). At this location, the recorded head's coordinate in camera space is $(X_z, Y_z, Z_z)_c$. A coordinate in camera space could then be converted to the work space using Equation 1 and 2. This calibration only needs to be performed once whenever the camera is moved to a new location. One can test the calibration result by moving in x, y, z axes and check the motion in both camera space and workspace as shown in Figure 3.

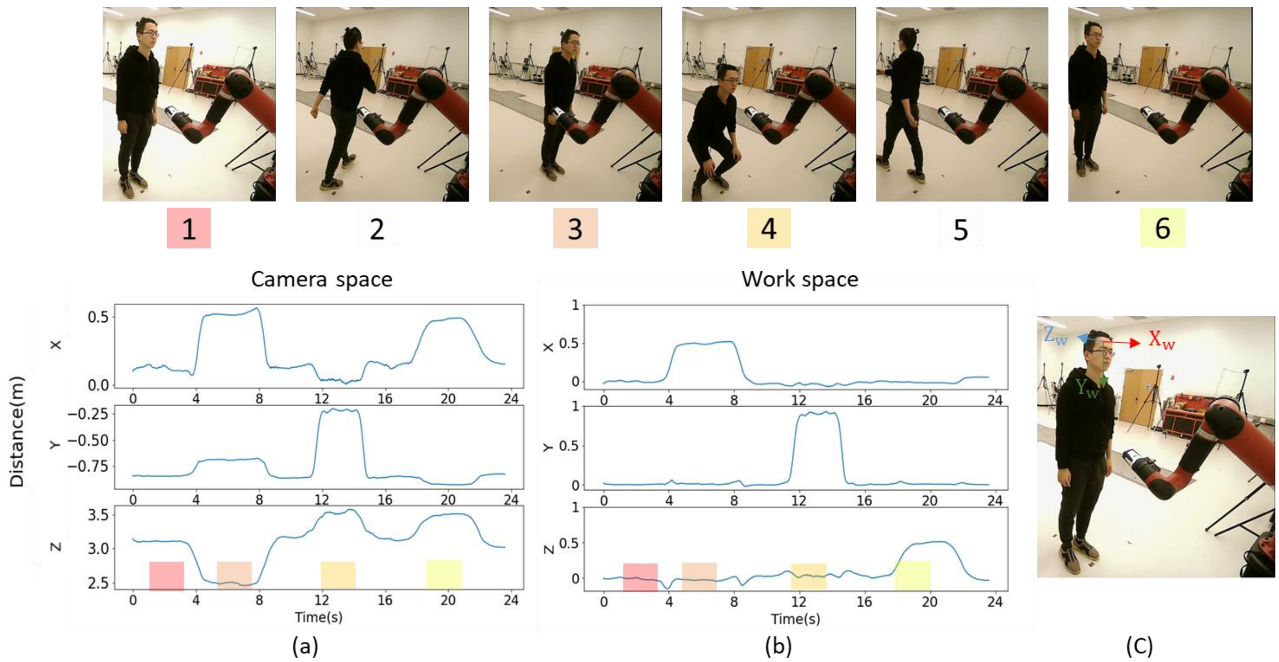

Figure 3. The video captured that one was moving in x, y, z directions of robot workspace (top frame 1-6). The recorded motion in the camera space (a) and workspace (b) are shown. The colors of the rectangles in (a) and (b) indicate specific timeframe from 1 to 6. in video and plot. The workspace coordinate system is shown in (c).

$$\mathbf{T} = \mathbf{S} \cdot \mathbf{R} = \begin{pmatrix} \frac{1}{L_x} & 0 & 0 \\ 0 & \frac{1}{L_y} & 0 \\ 0 & 0 & \frac{1}{L_z} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \qquad Eq. 1$$

$$\mathbf{P_w} = (\mathbf{P_c} - \mathbf{P_{0c}}) \cdot \mathbf{T}^{-1} \qquad Eq. 2$$

where

$$\boldsymbol{x} = (X_x - X_0, Y_x - Y_0, Z_x - Z_0)$$
$$\boldsymbol{y} = (X_y - X_0, Y_y - Y_0, Z_y - Z_0)$$
$$\boldsymbol{z} = (X_z - X_0, Y_z - Y_0, Z_z - Z_0)$$

$$\mathbf{R} = \begin{pmatrix} \frac{\boldsymbol{x}}{|\boldsymbol{x}|} \\ \frac{\boldsymbol{y}}{|\boldsymbol{y}|} \\ \frac{\boldsymbol{z}}{|\boldsymbol{z}|} \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} \frac{|\boldsymbol{x}|}{L_x} & 0 & 0 \\ 0 & \frac{|\boldsymbol{y}|}{L_y} & 0 \\ 0 & 0 & \frac{|\boldsymbol{z}|}{L_z} \end{pmatrix}$$

R is the approximate rotational matrix, **S** is the scaling matrix, $\mathbf{P_w}$ indicates the 3D coordinate of a point in workspace, and $\mathbf{P_c}$ indicates the 3D coordinate of the same point in camera space.

Collision avoidance

To actively avoid human-robot collision, the robot should be aware of worker's existence and actively respond to potential collisions. In this study, we consider two collision avoidance schemes. The first scheme is referred as *static collision avoidance,* where an unsafe zone is defined as a bounding cubic with a one-meter length in x-, y-, and z-direction with the center located at the end effector. If a worker's head position is within this unsafe zone, robot arm will retract. Please note that the actual shape and size of a real-world unsafe zone should be adjusted according to the configuration of the robot and layout of a workspace. The second scheme is referred as *dynamic collision avoidance*, which further considers workers' walking speed. Given a worker's current head position is $\mathbf{P}(T)$, the worker's head position after time t is $\mathbf{P}(T+t) = \mathbf{P}(T) + t \cdot \mathbf{V}(T)$, where $\mathbf{V}$ is the velocity vector at time T and is derived from the differential of position.

Function validation

In this study, we simulated a scenario where a worker first approached a co-robot and then left the co-robot in a different direction. One robot operator approached the co-robot twice (see Figure 4 caption for the details of operators moving path), once under the static scheme and once under the dynamic scheme. Whether a worker is in the unsafe zone is determined by calculating $\mathbf{P}(T)$ or $\mathbf{P}(T+t)$ in static/dynamic collision avoidance scheme, where t is set to one second.

The results indicated that the co-robot could actively retract its arm before the collision as the operator is approaching. The co-robot actions during static collision avoidance and dynamic collision avoidance are shown in Figure 4. For the dynamic collision avoidance, the co-robot responded to worker's approaching and leaving motion earlier than the static collision avoidance.

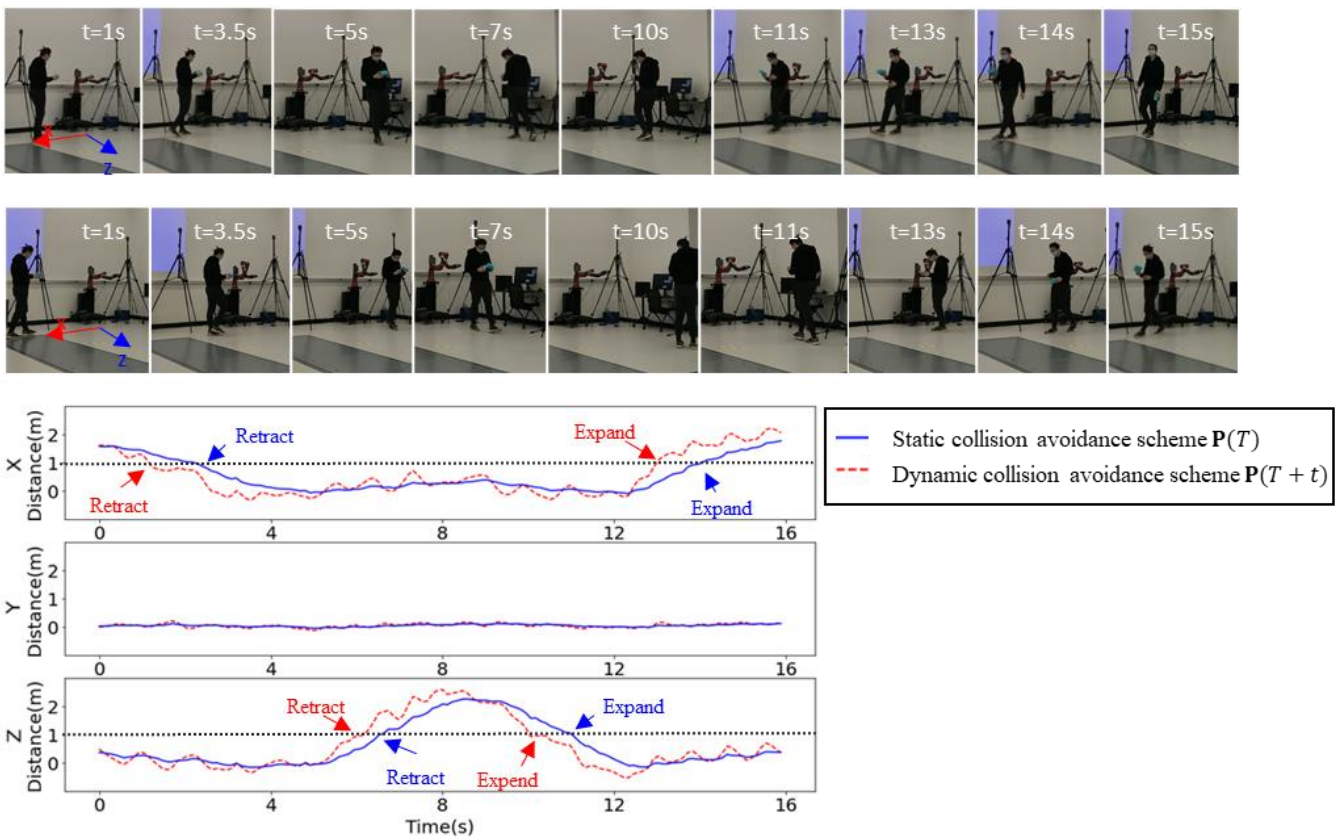

Figure 4. Preliminary validation of the proposed collision avoidance schemes. *Top*: A robot operator approached the co-robot in x-direction, turned right, and left in z-direction. This operator then turned back, approached the robot in –z-direction, turned left, and left the co-robot in x-direction. The first row of the key frames shows the co-robot's behavior when static collision avoidance scheme is adopted. The second row shows the co-robot's behavior when dynamic collision avoidance scheme is adopted. *Bottom*: The $\mathbf{P}(T)$ in the static collision avoidance scheme and $\mathbf{P}(T+t)$ in dynamic collision avoidance scheme are compared over time. The two curves are matched by the initial time instant of operator's move. Co-robot arm retraction and expansion time instants are marked. The co-robot responds earlier to the worker's approaching/leaving motion when dynamic collision avoidance scheme is applied.

DISCUSSION

This study proposed a collision avoidance method using a single RGB camera and computer vision algorithms. The preliminary validation indicates the proposed method is robust for allowing a co-robot to understand the existence of a surrounding worker and perform consequent collision avoidance actions.

It should be noted that while the dynamic collision avoidance scheme makes the co-robot to retract its end effector earlier as a worker is apporaching, this scheme also makes the co-robot to expand its end effector earlier as a worker is moving away. This may lead to a collision if a worker suddenly stops moving within the hazard zone, but the predict position of this worker is out of the hazard zone. Therefore, it might be beneficial to use dynamic scheme for co-robot arm retraction, and static scheme for co-robot arm expansion.

The predicted worker's position also showes a zig-zag behavior. This is because human walking speed is not a constant value due to human bipedal motion. The speed variance may incorrectly trigger robot actions. A low-pass filter could remove this zig-zag behavior, but a filter can also introduce phase shift which may lead to robot action delay. Therefore, the type and order of a filter need to be carefully selected.

There are a few limitations that need to be addressed. First, the optical features of a camera could influence the accuracy of the estimated workers' position. Specifically, some cameras could show distortions (e.g., a fisheye webcam), and these distortions could lead to less accurate reconstructed human pose. This limitation could be partially addressed by calibrating the camera calibration (Zhang, 2000). Second, the robot's retracting speed could be affected by the robot's joint motion speed. Thus, the parameter *t* in the dynamics collision avoidance method need to be carefully set to ensure the robot has enough time to retract its end effector. Third, in the current study we only harnessed workers' head position data for detecting potential collision. In future works, we will consider adopting the position of other body segments, such as the wrist joint, the location of which can be closer to the co-robot.

In a shell, the proposed method provides a potential simple solution for collision avoidance during human-robot collaboration. As this proposed method allows co-robot to understand workers' posture, it can be also applied in other applications, such as human-robot communication through workers' gestures (Abavisani et al., 2019; Li et al., 2019).

ACKNOWLEDGEMETNTS

This manuscript is supported by the National Science Foundation under Grant No 1822477 and 2024688.

REFERENCES

Abavisani, M., Joze, H. R. V., & Patel, V. M. (2019). Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1165–1174.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186.

Flenniken, W., Wall, J., & Bevly, D. (2005). Characterization of various IMU error sources and the effect on navigation performance. *Ion Gnss*, 967–978.

Furtado, J. S., Liu, H. H. T., Lai, G., Lacheray, H., & Desouza-Coelho, J. (2019). Comparative analysis of optitrack motion capture systems. In *Advances in Motion Sensing and Control for Robotic Applications* (pp. 15–31). Springer.

Girshick, Y. W. and A. K. and F. M. and W.-Y. L. and R. (2019). *Detectron2*.

Li, G., Tang, H., Sun, Y., Kong, J., Jiang, G., Jiang, D., Tao, B., Xu, S., & Liu, H. (2019). Hand gesture recognition based on convolution neural network. *Cluster Computing*, *22*(2), 2719–2729.

Matthias, B., Kock, S., Jerregard, H., Kallman, M., Lundberg, I., & Mellander, R. (2011). Safety of collaborative industrial robots: Certification possibilities for a collaborative assembly robot concept. *2011 IEEE International Symposium on Assembly and Manufacturing (ISAM)*, 1–6.

Michalos, G., Makris, S., Spiliotopoulos, J., Misios, I., Tsarouchi, P., & Chryssolouris, G. (2014). ROBO-PARTNER: Seamless human-robot cooperation for intelligent, flexible and safe operations in the assembly factories of the future. *Procedia CIRP*, *23*, 71–76.

Mohammed, A., Schmidt, B., & Wang, L. (2017). Active collision avoidance for human–robot collaboration driven by vision sensors. *International Journal of Computer Integrated Manufacturing*, *30*(9), 970–980.

Murashov, V., Hearl, F., & Howard, J. (2016). Working safely with robot workers: Recommendations for the new workplace. *Journal of Occupational and Environmental Hygiene*, *13*(3), D61–D71.

NIOSH. (2017). *NIOSH Presents: An Occupational Safety and Health Perspective on Robotics Applications in the Workplace.* Health, National Institute of Occupational Safety And.

Pavllo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2019). 3d human pose estimation in video with temporal convolutions and semi-supervised training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7753–7762.

Plantard, P., Auvinet, E., Pierres, A.-S. Le, & Multon, F. (2015). Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin. *Sensors*, *15*(1), 1785–1803.

Prayudi, I., & Kim, D. (2012). Design and implementation of IMU-based human arm motion capture system. *2012 IEEE International Conference on Mechatronics and Automation, ICMA 2012*, 670–675. https://doi.org/10.1109/ICMA.2012.6283221

Schäffer, A. A., Eiberger, O., Grebenstein, M., Haddadin, S., Ott, C., Wimböck, T., Wolf, S., & Hirzinger, G. (2008). Soft robotics, from torque feedback-controlled lightweight robots to intrinsically compliant systems. *IEEE Robotics and Automation Magazine*, *15*(3), 20–30.

Schlegl, T., Kröger, T., Gaschler, A., Khatib, O., & Zangl, H. (2013). Virtual whiskers—Highly responsive robot collision avoidance. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5373–5379.

Schmidt, B., & Wang, L. (2013). Contact-less and programming-less human-robot collaboration. *Procedia CIRP*, *7*, 545–550.

Sherwani, F., Asad, M. M., & Ibrahim, B. (2020). Collaborative robots and industrial revolution 4.0 (ir 4.0). *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, 1–5.

Xie, Z., Li, L., & Xu, X. (2021). Real-time driving distraction recognition through a wrist-mounted accelerometer. *Human Factors*, 0018720821995000.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334.