HAL
open science

Improving Human-Robot Collaboration in TV Assembly through Computational Ergonomics: Effective Task Delegation and Robot Adaptation

Brenda Elizabeth Olivas Padilla, Dimitris Papanagiotou, Gavriela Senteri, Sotiris Manitsaris, Alina Glushkova

To cite this version:

Brenda Elizabeth Olivas Padilla, Dimitris Papanagiotou, Gavriela Senteri, Sotiris Manitsaris, Alina Glushkova. Improving Human-Robot Collaboration in TV Assembly through Computational Ergonomics: Effective Task Delegation and Robot Adaptation. The 2023 IEEE Conference on Systems, Man, and Cybernetics, IEEE, Oct 2023, Oahu, United States. hal-04311425

HAL Id: hal-04311425

https://hal.science/hal-04311425v1

Submitted on 28 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Improving Human-Robot Collaboration in TV Assembly through Computational Ergonomics: Effective Task Delegation and Robot Adaptation

Brenda Elizabeth Olivas-Padilla
Centre for Robotics, Mines Paris
Université PSL, Paris, France
brenda.olivas@minesparis.psl.eu

Dimitris Papanagiotou
Centre for Robotics, Mines Paris
Université PSL, Paris, France
dimitris.papanagiotou@minesparis.psl.eu

Gavriela Senteri
Centre for Robotics, Mines Paris
Université PSL, Paris, France
gavriela.senteri@minesparis.psl.eu

Sotiris Manitsaris
Centre for Robotics, Mines Paris
Université PSL, Paris, France
sotiris.manitsaris@minesparis.psl.eu

Alina Glushkova
Centre for Robotics, Mines Paris
Université PSL, Paris, France
alina.glushkova@minesparis.psl.eu

***Abstract*—The high prevalence of work-related musculoskeletal disorders (WMSDs) could be addressed by optimizing Human-Robot Collaboration (HRC) frameworks. In this context, this paper proposes a methodology for ergonomically effective task delegation and HRC for manufacturing applications based on two hypotheses. The first hypothesis states that it is possible to rapidly quantify ergonomically professional tasks using motion data from a few wearable sensors and then delegate the high-risk tasks to a collaborative robot. The second hypothesis is that, compared to typical HRC frameworks involving physical interaction, the ergonomics and safety of an HRC scenario can be enhanced by combining gesture recognition and pose estimation. These remove unnecessary motions that could expose operators to ergonomic risks, decrease the amount of physical effort, and make the robot aware and responsive to the presence and gestures of the operators. The methodology is evaluated by optimizing the HRC scenario of a television manufacturing process, yet it is described how it can be reconfigured for other industrial scenarios. The effect of the temporal and spatial adaptation on the operator's range of motion was analyzed through three separate experiments. The effectiveness of HRC is measured through the standard key performance indicators (KPIs); however, to evaluate the collaboration and required physical demand, two KPIs are proposed in this paper. These are the rate of spatial adaptation and the rate of reduction in the operator's motion. The results demonstrated that the methodology enhanced the ergonomics and efficiency of the production process. First, the robot was delegated two tasks identified as the most ergonomically dangerous for human operators. Then the optimized HRC achieved an average rate of spatial adaptation of 29.37% and a decrease in operator movement of 28.8% across 14 subjects compared to HRC frameworks that do not include spatial and temporal adaptation.**

***Index Terms*—Human-Robot Collaboration, Wearable sensors, Ergonomics, Gesture recognition, Pose estimation, Assembly.**

The research leading to these results has received funding by the EU Horizon 2020 Research and Innovation Programme under grant agreement No. 820767, project CoLLaboratE.

I. Introduction

Industry 4.0 has resulted in a rise in research in the field of Human-Robot Collaboration (HRC). Robotic agents are therefore being integrated into the work routine, not to take the position of human operators but to assist them in accomplishing complicated and physically demanding tasks. However, it is critical to consider ergonomic factors to design HRC installations properly and prevent operators from developing work-related musculoskeletal disorders (WMSDs). Thus, current approaches for task delegation and collaboration between operators and robots must be improved by considering ergonomics to maximize operators' comfort and production efficiency in industrial co-production cells. With this in mind, a methodology for ergonomically effective task delegation and HRC is proposed in this paper. This methodology is founded on two hypotheses. The first hypothesis (H1) is that by utilizing a reduced amount of motion capture (MoCap) data from wearable inertial sensors, operators' movements can be accurately measured, allowing for a more thorough ergonomic analysis of their actions and facilitating task delegation when implementing HRC frameworks. The ability to record accurate measurements for ergonomic analysis is essential as it provides quantitative measures of operators' performance. The second hypothesis (H2) is that gesture recognition used for temporal adaptation and pose estimation for spatial adaptation enable a more natural and ergonomic HRC. The reason is that only motions that are convenient for the operators are performed, avoiding unnecessary movements that may expose human operators to ergonomic risks. Gesture recognition is one of the methods used to achieve contactless communication between a robot and a human operator. It can be defined as the process of transforming movements into a form that a machine can easily interpret. On the other hand, spatial adaptation is the process through which the robot adjusts its movements according to

the anthropometric characteristics of each operator. Including human pose estimation enriches the robot's perception and enables its spatial adaptation in the HRC. Thus, the robot can assist operators in reducing their range of motion and physical effort necessary to fulfill their professional duties. Fig. 1 illustrates the 2-step methodology proposed in this paper for ergonomically optimizing an industrial HRC scenario. First, wearable inertial sensors are used to conduct an in-depth analysis and ergonomic evaluation of professional tasks for task delegation. The proposed HRC then employs an RGB camera in egocentric vision and a frontal RGB-D camera for real-time gesture recognition and pose estimation, respectively. The methodology is evaluated based on its application to an actual television (TV) manufacturing process. Following, Section II discusses state of the art in ergonomic analysis using MoCap data and HRC frameworks. Section III explains the methodology for ergonomically evaluating professional tasks and the results obtained. Next, Section IV details the HRC framework and the calculated KPIs. Finally, Section V presents the conclusion and future work.

II. State of the Art

The activities performed by operators in the industrial sector are becoming more challenging and complex in order to meet market demands within certain time limits, job specifications, and budget constraints. Operators must go beyond their natural physical limitations to undertake repetitive jobs for long periods to complete the required tasks. Being subjected to such constant physical strain leads to WMSDs. Ergonomists have developed a variety of methods for evaluating work-related tasks. The methods based their analysis on theoretical knowledge of human physical limitations and abilities indicated by known standards (e.g., ISO 11226:2000 and EN 1005-4). Popular methods are the Rapid Upper Limb Assessment (RULA) [1], Rapid Entire Body Assessment (REBA) [2], and European Assembly Worksheet (EAWS) [3]. To implement these methodologies, the ergonomist observes an operator executing the task under evaluation and annotates various body part postures on a worksheet to estimate an score. Yet, because these approaches rely on the ergonomist's perception and experience, scoring can be subjective and have inter-variability. Previous studies have used biomechanical simulations to compute ergonomic metrics (posture, physical effort, and energy spent) [4], [5]. Nevertheless, adopting these analyses in industrial applications that demand rapid reconfigurability has been hindered due to several factors [5], [6], [7]. These include requiring specialized and costly equipment, frequent calibration procedures, limited capture volumes, and substantial installation and operational expenses. Thus, there is still a need for methodologies that employ MoCap technologies that are practical to implement in real-world settings and can accurately estimate the ergonomic risk of any representative set of manipulation actions performed in industry.

Collaborative robots are becoming increasingly common as an automated solution for making workplaces more ergonomic, cost-effective, and flexible. As a result, early research focused on creating HRC frameworks that physically couple humans and robots. For example, some applications are for co-carrying [8], [9] and co-manipulation [10]. However, most existing real-time HRC frameworks are developed around the concept of human avoidance to assure safety, i.e., preventing unintentional accidents between humans and robots [11], [12]. A study that directly addressed human ergonomic demands was done by Bestick et al. [13], who investigated the most convenient robot configurations when giving over an object to a human operator through offline biomechanical simulations. A priori assumptions were made regarding the cost function based on the distance of human joints from the neutral position. Lorenzini et al. [14] and Kim et al. [15] used overloading joint torque as a cost in the offline simulation to optimize the human body configuration in co-carrying and co-manipulation activities based on the human body stability, task, and workspace conditions. Now, the primary focus of ergonomics research lies in incorporating and applying evaluation methods effectively within work environments without requiring offline biomechanical analysis in laboratory settings. Therefore, it is crucial to explore the potential use of user-friendly MoCap technologies that are comfortable and do not disrupt daily work activities. Also, future research should concentrate on designing methods for distributing roles in HRC frameworks and others that facilitate online planning and adjustment of robot movements. This online planning and robot adaptation can be evaluated based on Key Performance Indicators (KPIs) that consider ergonomic risk factors. Previous research provided metrics that primarily assessed the productivity of collaborative robots [16], [17]. For instance, neglect tolerance (NT), concerned with how long a human can ignore a robot, has been used to measure the robot's effectiveness in interacting with its operators. Another standard indicator is the interaction effort (IE) which denotes the time necessary to engage with the robot. The combination of NT and IE produces the robot attention demand (RAD), a dimensionless metric that quantifies the level of focus that the robot requires from the operator. Finally, there is also the measurement of the efficiency of the robot (ER), which involves the time it takes for the robot to move in relation to the total time of the whole routine. Consequently, to measure the job quality of operators in HRC frameworks online, rather than just their productivity performance, new KPIs that consider human factors must be developed.

To summarize, practical and rapidly reconfigurable frameworks with real-time data processing capabilities are necessary to handle the industry's ergonomics challenges while ensuring productivity. This paper presents a methodology for creating an ergonomic and reconfigurable HRC framework, following up on the authors' prior work while developing natural HRC [18] and analyzing postural risk factors [19].

III. Automatic ergonomic evaluation

This section describes the task delegation approach in which real professional tasks are automatically evaluated using a minimal set of inertial sensors.

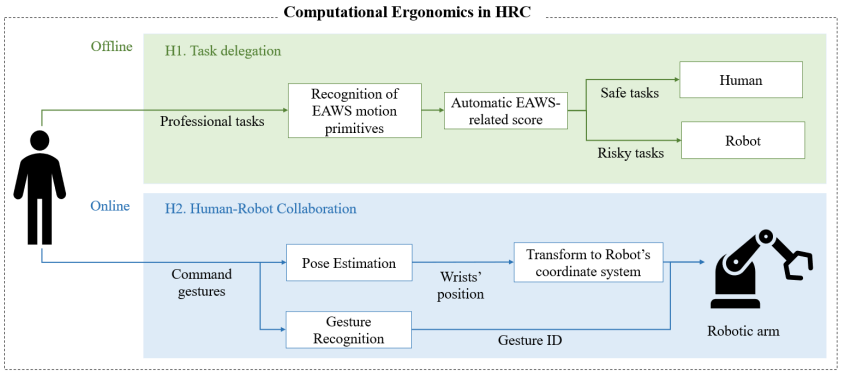


Fig. 1: Pipeline for ergonomically optimizing industrial co-production cells with HRC.

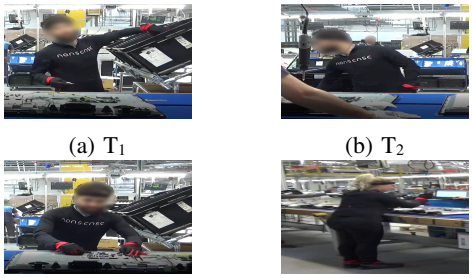
A. Motion capturing of the television assembly scenario

For the initial study and recording of tasks performed on a TV production line, the BioMed bundle motion capture system from Nansense Inc. (Baranger Studios, Los Angeles, CA, USA) was used. This system consisted of a full-body suit composed of 52 IMUs strategically placed throughout the body and hands. At a rate of 90 frames per second, the IMUs were used to calculate the Euler joint angles of body segments on the articulated spine chain, shoulders, limbs, and fingertips. Wearable sensors were chosen for the ergonomic evaluation because they provide more precise measurements of the operators' posture than optical MoCap systems such as RGB cameras.

Two healthy adults, one male and one female, wore the IMU-based suit during the MoCap recording session. The recording took place on a real production line over the course of an eight-hour shift, with the operators placing circuit cards into the frame of TVs. The entire assembly procedure can be broken down into four main tasks, each illustrated in Fig. 2. The first task is grabbing a circuit card from a container (T_1); the second is taking a wire from a second container (T_2); the third involves connecting the card and wire and placing them on the TV frame (T_3); the fourth task corresponds to drilling the cards on the TV frame (T_4). The whole assembly was recorded 108 times. Following the recordings, the MoCap data was pre-processed offline before analysis. Low pass filtering with a Butterworth filter with a cut-off frequency of 5hz was used to reduce noise, and the common zero velocity update algorithm was used to eliminate drift that occurred in some iterations by electromagnetic interference. When the algorithm detects a period of zero or near-zero velocity, it updates the orientation and position of the subject to reduce accumulated errors. The full details of the processing of the recordings are given in [20].

B. Modeling of motion primitives for ergonomic evaluation

An automatic posture evaluation system was built to identify potentially hazardous postures in professional tasks. The system is comprised of Hidden Markov Models (HMMs) that learned to recognize motion primitives. The motion primitives consist of small sequences of postures with varying ergonomic



(a) T_1 (b) T_2 (c) T_3 (d) T_4

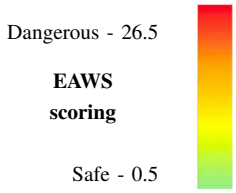
Fig. 2: Professional tasks for TV assembly.

risk levels, according to EAWS evaluation. Based on the detected motion, an ergonomic score is computed. For generating the dataset with the motion primitives, MoCap recordings were made in a laboratory. The neutral environment sessions involved ten healthy individuals, three females and seven males. None of them suffered from any musculoskeletal injury. The protocol contained 28 different motion primitives, each recorded three times. The motions can be classified into three broad categories: those performed standing, those performed while seated in a chair, and those performed while kneeling. The small movements ($\approx$ 4 seconds) progressed from comfortable to increasingly uncomfortable postures (raising arms, bending forward, rotating the torso). Each motion has assigned an EAWS risk score, ranging from 0.5 to 26.5. The higher the ergonomic risk score, the greater the risk. In prior research of these motion primitives, their spatiotemporal dynamics were analyzed using a Gestural Operational Model (GOM) [19]. As a result, it was discovered that high recognition accuracy can be achieved with only five inertial sensors. These sensors are positioned on the lumbar spine, the left upper arm, the right shoulder, the right upper leg, and the left forearm. Consequently, the proposed system for evaluating professional tasks only requires measurements from these five sensors.

C. Automatic postural evaluation of professional tasks

To conduct the ergonomic evaluation, the recordings of the professional tasks are segmented into short windows of similar duration to the motion primitives and then provided to the

TABLE I: Summary statistics of the EAWS scores.



Tasks	EAWS scores		
	Mean	σ	Mode
T_1	16.02	2.65	17.50
T_2	15.02	3.43	16.00
T_3	10.76	3.68	8.50
T_4	11.50	3.19	12.50

automatic evaluation system. The resulting likelihoods indicate the detected motion primitive. All motion primitives identified throughout each activity are annotated and utilized to calculate each task's mode and average ergonomic score. The score is determined using EAWS's score calculations [3], and the statistics are used to decide which tasks expose operators to the higher risk.

D. Results and discussion of the ergonomic evaluation

Before segmenting professional tasks into motion primitives, the 28 HMMs were trained using a left-right topology with seven internal states that yielded the maximum recognition performance. Their performance was evaluated using the all-shots approach and achieved an overall F-score of 95.03%, precision 94.97%, and recall 95.30%. The assembly tasks were then evaluated with the automatic evaluation system, resulting in the EAWS scores illustrated in Table I. Table I contains the mean, standard deviation, and mode of the scores calculated for each task.

According to these results, the majority of iterations of tasks T_1 and T_2 can be classified as medium-risk motions, while iterations of tasks T_3 and T_4 are classified as low-risk motions. For T_1, the motion primitives detected most frequently resembled motions in which the elbows are raised above shoulder level while the torso is laterally bent. These results are expected as operators must rotate and laterally bend their torsos to retrieve a circuit card from a container. Due to the container's location, operators must also raise their arms above shoulder level, as illustrated in Fig. 2(a). The motion primitives detected for T_2 corresponded more to motions where there is both bending and rotation of the torso and stretching of the arms. These results match T_2, where the operators bend to pick up a wire from a container and then connect it to the card. The motion primitives detected for T_3 and T_4 were primarily torso rotations and working with arms bent around 90°. These motion primitives are visible in Fig 2(c) and 2(d). In these tasks, operators were only required to slightly rotate their torso due to the constant movement of the TV frame caused by the conveyor belt, but they did not need to stretch their arms or strongly bend forward to reach the TV frame. Since T_1 and T_2 involve assuming awkward postures such as rotating the torso while bending forward or raising the arms above shoulder level, they represent a major ergonomic risk than T_3 and T_4. EAWS recommends that, if possible, tasks with moderate risk must be redesigned; otherwise, the risk must be controlled through other means. Therefore, it is proposed to delegate T_1 and T_2 to the robot and leave T_3 and T_4 to the operators. These results were compared to those calculated by the factory's ergonomist, who manually calculated the ergonomic risk of each task using the Rapid Entire Body Assessment (REBA) worksheet [2]. According to the results of this study, this postural evaluation categorized T_1 and T_2 as medium-risk tasks with a REBA score of 7, which suggests redesigning the task to improve ergonomics. On the other hand, T_3 and T_4 were given a score of 2, meaning movements of low risk.

Consequently, the proposed task delegation approach that automatically detects postural risk factors utilizing a reduced set of inertial sensors was validated for this workplace scenario. This analysis can be applied to other professional tasks for rapid reconfigurability, just as it was done with TV assembly tasks. First, the professional tasks to be evaluated need to be recorded with the designated inertial sensors and placed according to the standards of the International Society of Biomechanics (ISB). Next, calculate the Euler angles of each body part measured by each sensor. Then, segment the data of the tasks into four-second windows, ideally with a two-second overlap to cover the entire task. Finally, apply the automatic postural evaluation to the segmented tasks, which would indicate the motion primitives detected and the estimated EAWS score. Depending on the nature of the identified high-risk tasks, it is determined whether they should be delegated to a collaborative robot or the production cell should be redesigned. Additionally, the proposed approach for task delegation can serve for high-frequency monitoring of the operators, providing human ergonomists with quantitative measures of the workers' performance and postural risk detected during the shift.

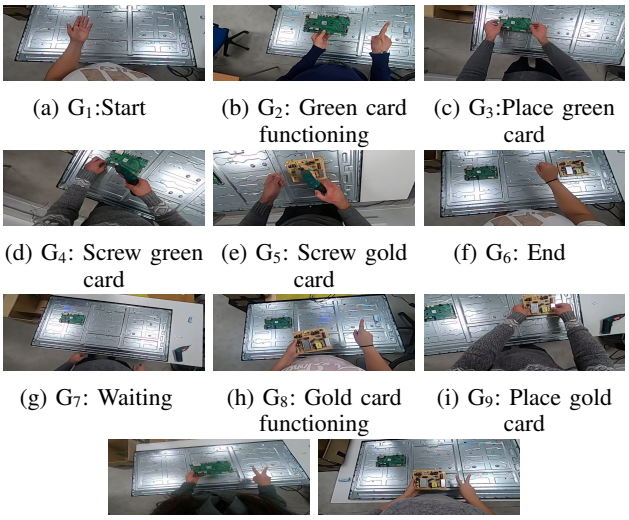
IV. Optimization of work-space scenario

Based on the results of the automatic ergonomic evaluation, this section describes the integration of a collaborative robot into the manufacturing process. To improve the interaction of the collaborative robot and the operator and reduce the operator's range of motion, gesture recognition and pose estimation modules were included. Three experiments were conducted using the TV assembly scenario to evaluate the proposed HRC framework using standard KPIs and two new KPIs proposed in this paper.

A. Gesture recognition with 3DCNNs

According to the task delegation results, the assembling routine is first divided into sub-tasks performed by the robot or operators. 3D Convolutional Neural Networks (3DCNNs) are used for the gesture recognition module in real-time, which establishes a gestural communication between the operator and the robot. As proven in recent research [18], gesture recognition decreases production time in a real-life industrial scenario and ensures workers use ergonomic postures.

3DCNNs were chosen since they are an end-to-end method that can get as input RGB images without the prior extraction of features like the case is for other methods, such as the Hidden-Markov models. These networks were trained using



(a) G_1:Start (b) G_2: Green card functioning (c) G_3:Place green card

(d) G_4: Screw green card (e) G_5: Screw gold card (f) G_6: End

(g) G_7: Waiting (h) G_8: Gold card functioning (i) G_9: Place gold card

(j) G_{10}: Green card not functioning (k) G_{11}:Gold card not functioning

Fig. 3: Command gestures of the gesture recognition module.

a dataset consisting of command gestures adapted for TV assembly and recorded with an RGB camera in an egocentric view. The challenge with using egocentric data for recognition is that the head and hands move in parallel but do not always follow each other. Even if the hands are a significant part of this module, they can be prominent within the frame but can also be partly or even totally out of the main view. Nevertheless, the added value of an egocentric perspective is that the network will provide gesture recognition results independent of the users' anthropometric characteristics and position in the environment.

To train the proposed network, a dataset was first created from this particular HRC scenario, consisting of RGB images recorded with a GoPro Black, placed with a headband on the head of the human operator. The recording was made at an 848x480 resolution and 20 frames per second. Concerning the architecture of the 3DCNNs, it consisted of 6 convolution layers, with max-pooling and batch normalization in between. The Adam optimizer was used with a learning rate of 0.000001. To facilitate the real-time HRC, a sliding window of 5 frames was used, as it was the one that gave the best results among those of 10, 15, and 20 frames. For the creation of the TV-assembly dataset, a group of 14 operators, four females and ten males, were recorded performing six gestures and five postures. Thus, there are 11 classes in total, each corresponding to a unique command for the collaborative robot, presented in Fig. 3. The objects involved are a TV frame and two circuit cards: the power supply (PSU-gold card) and the mainboard (chassis-green card). These gestures were adapted for TV assembly, but they can be customized for other scenarios by modifying the actions performed by the robot.

B. HRC for television manufacturing

For the proposed HRC scenario, the UR3 robotic arm (Universal Robots, Odense, Denmark)is used. The external parts that are used for grasping and releasing the circuit cards are from ROBOTIQ (gripper: 2F-140 & force torque sensor: FT-300-S). The Robot Operating System (ROS) is used to control the robotic arm. Through a UDP communication protocol, the robotic arm receives real-time ID messages from the gesture recognition module. In the new routine, the operator performs the start gesture (G_1) to notify the robot that the assembly routine starts. Next, the robot approaches the card container, retrieves the initial green card, and hands it to the operator in the pre-defined handover position. By pressing the Force sensor, the operator releases the card and verifies its functionality. If the card is functional, the operator performs G_2 and places it on the TV frame (G_3), while the robotic agent moves toward the card box to retrieve the gold card. If the green card is not functional, the operator executes G_{10} to notify the robot, which then brings a replacement green card. When the robotic arm delivers a functional card, the operator performs G_2 and then places and screws the card on the TV frame (G_3 and G_4). This procedure is repeated until both the green and gold cards are placed on the TV frame. Then, the human operator executes G_{11} to signal the routine's completion. Safety regulations for collaborative robotics are followed according to standards (ISO 10218 and TS 15066). The gesture recognition module is also designed with dual-level control to ensure the operator's safety in the HRC scenario. More precisely, the robot receives the ID of the recognized gesture only after it has been correctly recognized for 20 consecutive time frames. This may raise concerns about any latency observed during the TV assembly process. However, it was observed that the time interval between capturing a frame and its correct recognition was between 0 and 0.08 ms, indicating no significant latency. The second level of control was performed within ROS, where the robot accepted only gestural IDs that corresponded to the assembly workflow as defined by the routine. As a result, if an unintended incident occurs, the operator retains full control as the robotic arm strictly adheres to the routine.

The introduction of the robotic arm and a gesture recognition module requires only minor torso rotations from the human operator. To facilitate natural collaboration and assist the operator in performing only ergonomically safe motions, a posture estimation module is added that enables the robot to spatially adapt to the operator. The methodology for pose estimation and spatial adaptation of the robot is described in detail in an earlier publication by the authors at [18]. The spatial adaptation refers to the fact that the robotic arm does not place the cards in a fixed position but rather adapts to the operator's anthropometric characteristics, thereby improving the operator's posture. An Intel-RealSense RGB-D camera is utilized to capture the operator's upper body for pose estimation and the robot's spatial adaptation. The X and Y axes of the camera are first aligned with the X and Z axes of the robot. Then, the operator's skeleton and the position and velocity of the operator's wrists are extracted. The spatial adaptation then consists of the robotic arm tracking the operator's hand to approach it with a card when requested

TABLE II: Standard KPIs measured for each experiment.

Experiments	Neglect tolerance	Interaction effort	RAD	Efficiency of the robot
1. Physical interaction	97.47 (σ: 4.60)	13.53 (σ: 2.80)	0.12 (σ: 0.001)	25.90*
2. Gesture recognition	60.14 (σ: 4.20)	21.90 (σ: 3.10)	0.27 (σ: 0.001)	34.90*
3. Gesture recognition and pose estimation	76.73 (σ: 2.40)	28.33 (σ: 3.20)	0.26 (σ: 0.001)	30.85 (σ: 0.30)

TABLE III: Proposed KPIs to measure the collaboration of the robot with each operator.

KPI	Operator													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
SA (%)	39.10	33.30	21.10	27.50	30.40	31.90	27.10	31.80	13.40	33.90	43.50	32.10	18.70	27.40
RiOM (%)	31.40	33.10	24.40	27.10	32.10	27.30	24.50	26.80	37.40	20.60	45.90	20.80	21.30	24.50

by the operator and when the hand is motionless and in an accessible position for the robot.

C. Key performance indicators

To evaluate the HRC framework, NT, IE, RAD, and ER were calculated, which essentially assess the interaction between the robot and the operator as well as their productivity in the HRC. This paper introduces two new KPIs to evaluate the physical demand required of the operator in the HRC. The first refers to the percentage of robot spatial adaptation (SA) by adding pose estimation, and the second is for determining how much effort from the operator is reduced due to the gesture recognition module; this KPI is designated as the reduction in the operator's motion effort ($RiOM$). The KPI of robot spatial adaptation represents the ratio of the distance covered by the robot without spatial adaptation to the distance covered when the robot adjusts to the operator-specified position. The following formula is used to determine this KPI, where SA is the spatial adaptation percentage:

$$SA(\%) = \frac{\|AHP - WP\| - \|PHP - WP\|}{\|PHP - WP\|} \quad (1)$$

By measuring the position of the robot's ending point, it is obtained the waiting point WP, which is the robot's position when waiting for the instruction of the operator to provide a circuit card. The adapted handover position AHP is the position where the robot gives the circuit card according to the operator's hand, using pose estimation. Finally, the pre-defined handover position PHP is the position the robot leaves the circuit card without pose estimation, fixed for all operators. Centimeters are used to measure distances. The higher the rate of adaptation, the more effort the operator had to put in to reach the circuit card during the HRC scenario when there was no spatial adaptation of the robot.

The second KPI quantifies the difference in operators' motion before and after introducing gesture recognition. This KPI is calculated as follows:

$$RiOM(\%) = \frac{\|MwoGR\| - \|MwGR\|}{\|MwoGR\|} \quad (2)$$

The position of the operator's wrist is measured to calculate the distance covered by the wrist while interacting with the robot. $MwoGR$ is the distance covered by the wrist when there is no gesture recognition and the operator need to press the robot's force-torque sensor to continue the routine. $MwGR$ is the distance covered by the wrist when there is gesture recognition and the operators only perform a gesture with their hand to continue their routine. The higher the percentage of reduction in distance, the greater the convenience for the user in completing their routine with fewer physical movements. Consequently, $RiOM$ indicates the amount of effort reduced by the operator as a result of gesture recognition.

D. Experimental results

In order to evaluate the HRC scenario in terms of collaboration and operator performance, 14 operators, who did not participate in the creation of the training dataset, were recorded performing the proposed routine in three separate experiments. The 14 subjects consisted of seven males and seven females, either right- or left-handed, ranging in height from 1.60 to 1.90 meters.

The 3DCNNs were trained with the command gesture dataset using the 80:20 approach for the gesture recognition module. An F-score of 97.20% was achieved in recognizing the 11 gestures after 40 epochs and 98.60% by using early-stopping. For real-time recognition, the module presented a delay of 0.8 seconds. The confusion matrix is illustrated in Fig. 4, where it can be observed that the main false results are among G_4, which is the gesture where the human operator screws the green card on top of the TV frame, and G_7, where the operator waits for the robot for the next card. The main reason is the confusion caused by part of the frames from G_4 where the operator had the hands down as in G_7. When the algorithm was deployed in real-time, for the collaboration of the human operator with the robot, the challenge in the recognition results was noticed between gestures G_2 and G_{10}, as well as G_8 and G_{11}. The primary difficulty arises from the fact that the TV frame contains edges that can be mistaken for the fingers of the human operator. Additionally, it was observed that in real-time scenarios, lightning conditions affect the recognition results for the gestures mentioned above due to the reflection of the light on the metallic TV frame. Nevertheless, despite these difficulties, all 14 new subjects successfully completed the collaboration procedure, indicating that the accuracy of the gesture recognition module is sufficient even for users that were not part of the training dataset.

The case of failed recognition was handled in two ways. Firstly, the TV assembly routine consists of a specific se-

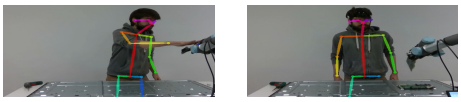
Actual Class \ Predicted Class	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}	G_{11}
G_1	46 100%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
G_2	0 0%	61 98%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
G_3	0 0%	0 0%	186 100%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
G_4	0 0%	1 2%	0 0%	265 75%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
G_5	0 0%	0 0%	0 0%	0 0%	847 100%	0 0%	0 0%	1 2%	0 0%	0 0%	0 0%
G_6	0 0%	0 0%	0 0%	0 0%	0 0%	59 100%	0 0%	0 0%	0 0%	0 0%	0 0%
G_7	0 0%	0 0%	0 0%	88 25%	0 0%	0 0%	340 100%	0 0%	0 0%	0 0%	0 0%
G_8	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	46 98%	0 0%	0 0%	1 2%
G_9	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	109 100%	0 0%	0 0%
G_{10}	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	42 100%	0 0%
G_{11}	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	55 98%

Fig. 4: Confusion matrix of the multi-class gesture recognition.

quence of gestures, and every class can be followed only by one or two at most other classes. The collaborative robot ignores recognition errors that do not follow the TV assembly procedure. Second, suppose an incorrect gesture is identified but is part of the normal assembly sequence. In that case, the collaborative robot only considers recognition results of 100% for 30 consecutive frames, which takes one second to compute.

Three experiments were conducted to assess the adaptation of the robotic arm and the operator's motion during each. In the first experiment, gesture recognition and spatial adaptation were disabled. Hence, operators were required to interrupt their routine and inform the robotic arm of their current action by pressing its force-torque sensor (vertically in order to proceed to the next step or horizontally if the received card is not functional). For the second experiment, only the gesture recognition module was enabled, so the operators received the circuit cards from a PHP. Finally, in the third experiment, gesture recognition and spatial adaptation were enabled and continuously provided information about the operators' actions to the robotic arm.

The KPIs NT, RAD, IE, and ER were calculated for each experiment to evaluate the different HRC frameworks and are given in Table II. However, for the proposed KPIs, SA is primarily used to compare the first experiment (physical interaction) with the third experiment (gesture recognition and spatial adaptation with pose estimation), whereas $RiOM$ is used to compare the first experiment with the second experiment (only gesture recognition). The measured SA and $RiOM$ for each operator are shown in Table III. Note that the greater the percentage of SA, the more difficult it was for the operator to receive the cards without spatial adaptation enabled. This is because the PHP was not possible to be convenient for every operator. For $RiOM$, the larger the percentage, the better, as it indicates a greater reduction in operator motion when gesture recognition is utilized.



(a) MwoGR: Press robot to start

(b) MwGR: The robot recognizes G_7 to start

Fig. 5: TV assembly with and without gesture recognition.

E. Discussion

The four standard KPIs are designed to assess the effectiveness of the performance of a robot in an HRC scenario. According to TableII, a larger value of NT indicates that less information is being conveyed to the robot about the operator's actions and intentions, which is evident in the experiment where only physical interaction takes place. However, this limitation can be addressed by incorporating gesture recognition and pose estimation modalities. RAD is also displayed and reflects the robot's ability to understand and adapt to the operator. The implementation of gesture recognition modality enhances the RAD because the operator is able to interact with the robot not only explicitly but also implicitly through the use of gestures. The efficiency of the robot is determined by calculating the percentage of time it spends in motion while executing a routine. In the first and second experiments, the handover position is pre-defined, and the duration of the motion of the robot remains constant throughout the experiment; this is denoted by an asterisk (*) in TableII. In this case, where the efficiency falls below 50 percent, it indicates that a single robotic arm can be utilized to simultaneously handle two different product lines.

These standard metrics do not take into account the potential benefits of different modalities of HRC at the operator's convenience. To address this gap, this paper proposes the use of two new movement-based KPIs (SA and $RiOM$). As stated previously, collaborative robotics can improve operators' working conditions and specifically contribute to preventing WMSD. According to the results achieved using the proposed HRC with gesture recognition and spatial adaptation, Table III illustrated a consistent decrease in the physical effort required of the operator. The spatial adaptation had the most impact on subject 11, who required more physical effort than the other subjects to reach the PHP. However, for subject 9, the difference in effort between using the PHP or AHP was the smallest, with a SA of 13.40%. Regarding $RiOM$, the gesture recognition module helped reduce motion the greatest for subject 9 and reduced motion by at least 20% for all subjects. This reduction can be illustrated in Fig. 5, where in Fig. 5(a), the operator was required to rotate his torso in order to touch the robot's sensor, which is located outside the TV frame. In Fig. 5(b), the robot recognizes when the operator has completed his task and can proceed to the next action. The average rate of spatial adaptation and reduction in operator motion was 29.37% and 28.37%, respectively, among the 14 subjects.

These results proved hypothesis H2 as the collaborative

robot can adapt to human factors and aids in improving ergonomics and efficiency, evidenced by the KPIs. Since the collaborative robot would be assigned to the most dangerous tasks, similar results could be expected for other assembly processes. The operator would only need to perform command gestures and safe tasks during the operation. The proposed KPIs revealed motion variations across experiments, demonstrating that they can be a relevant indicator to include in the evaluation for measuring how physically demanding and comfortable the installation is for the operators.

V. Conclusion and Future Work

This paper presents a methodology for task delegation and an HRC framework that maximize ergonomics and production efficiency in a television co-production cell. For this work, professional tasks performed on a television production line were captured using a small set of wearable sensors. Then, motion primitives with known ergonomic risks were detected across the entire professional tasks to estimate the tasks' ergonomic scores and identify the most dangerous ones. An optimized HRC was then proposed, in which hazardous tasks were delegated to the collaborative robot. The HRC scenario was enhanced by applying gesture recognition for temporal adaptation and pose estimation for spatial adaptation that allowed the human operator to collaborate with the robot using gestures while avoiding unnecessary movements that could cause physical strain. Lastly, two novel KPIs are proposed for measuring the physical demand required of operators when collaborating with a robot. These can complement the evaluation of HRC frameworks by considering not only the production efficiency but also the ergonomics and job quality of the operator.

Future research will focus on developing a frame-by-frame risk prediction method. The system would be capable of predicting future postures and their ergonomic risk. This type of system may be advantageous for HRC for various reasons. For example, the robot could provide preventive feedback to operators or protect them from risks by actively assisting them in adopting safer postures before and during a task.

Acknowledgment

The authors like to express their gratitude to Arçelik A.S. for their support in this work.

References

[1] L. McAtamney and N. Corlett, "RULA: A survey method for the investigation of work-related upper limb disorders," *Applied Ergonomics*, vol. 24, no. 2, pp. 91–99, 1993.

[2] S. Hignett and L. McAtamney, "Rapid Entire Body Assessment (REBA)," *Applied Ergonomics*, vol. 31, no. 2, pp. 201–205, apr 2000. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003687099000393

[3] K. Schaub, G. Caragnano, B. Britzke, and R. Bruder, "The European Assembly Worksheet," *Theoretical Issues in Ergonomics Science*, vol. 14, no. 6, pp. 616–639, 2013.

[4] W. Kim, J. Lee, L. Peternel, N. Tsagarakis, and A. Ajoudani, "Anticipatory robot assistance for the prevention of human static joint overloading in human–robot collaboration," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 68–75, 2018.

[5] A. Muller, C. Pontonnier, X. Robert-Lachaine, G. Dumont, and A. Plamondon, "Motion-based prediction of external forces and moments and back loading during manual material handling tasks," *Applied Ergonomics*, vol. 82, p. 102935, 2020. [Online]. Available: https://doi.org/10.1016/j.apergo.2019.102935

[6] M. Menolotto, D.-S. Komaris, S. Tedesco, B. O'Flynn, and M. Walsh, "Motion Capture Technology in Industrial Applications: A Systematic Review," *Sensors*, vol. 20, no. 19, p. 5687, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/19/5687

[7] R. M. Kanko, E. K. Laende, E. M. Davis, W. S. Selbie, and K. J. Deluzio, "Concurrent assessment of gait kinematics using marker-based and markerless motion capture," *Journal of Biomechanics*, vol. 127, p. 110665, 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0021929021004346

[8] X. Yu, W. He, Q. Li, Y. Li, and B. Li, "Human-Robot Co-Carrying Using Visual and Force Sensing," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 9, pp. 8657–8666, 2021.

[9] D. Sirintuna, A. Giammarino, and A. Ajoudani, "Human-Robot Collaborative Carrying of Objects with Unknown Deformation Characteristics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. abs/2201.1. IEEE, oct 2022, pp. 10 681–10 687. [Online]. Available: https://arxiv.org/abs/2201.10392http://arxiv.org/abs/2201.10392https://ieeexplore.ieee.org/document/9981948/

[10] A. Al-yacoub, Y. C. Zhao, W. Eaton, Y. M. Goh, and N. Lohse, "Robotics and Computer-Integrated Manufacturing Improving human robot collaboration through Force / Torque based learning for object manipulation," *Robotics and Computer-Integrated Manufacturing*, vol. 69, no. December 2020, p. 102111, 2021. [Online]. Available: https://doi.org/10.1016/j.rcim.2020.102111

[11] A. N. Sharkawy, P. N. Koustoumpardis, and N. Aspragathos, "Human–robot collisions detection for safe human–robot interaction using one multi-input–output neural network," *Soft Computing*, vol. 24, no. 9, pp. 6687–6719, 2020. [Online]. Available: https://doi.org/10.1007/s00500-019-04306-7

[12] M. El-Shamouty, X. Wu, S. Yang, M. Albus, and M. F. Huber, "Towards Safe Human-Robot Collaboration Using Deep Reinforcement Learning," in *Proceedings - IEEE International Conference on Robotics and Automation*. Paris, France: IEEE, 2020, pp. 4899–4905.

[13] A. Bestick, R. Bajcsy, and A. D. Dragan, "Implicitly Assisting Humans to Choose Good Grasps in Robot to Human Handovers," 2017, pp. 341–354. [Online]. Available: http://link.springer.com/10.1007/978-3-319-50115-4_30

[14] M. Lorenzini, W. Kim, E. D. Momi, and A. Ajoudani, "A New Overloading Fatigue Model for Ergonomic Risk Assessment with Application to Human-Robot Collaboration," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1962–1968. [Online]. Available: https://ieeexplore.ieee.org/document/8794044/

[15] W. Kim, L. Peternel, M. Lorenzini, J. Babič, and A. Ajoudani, "A Human-Robot Collaboration Framework for Improving Ergonomics During Dexterous Operation of Power Tools," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102084, 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0736584520302945

[16] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06*, vol. 2006. New York, New York, USA: ACM Press, 2006, p. 33. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1121241.1121249

[17] J. Crandall, M. Goodrich, D. Olsen, and C. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, no. 4, pp. 438–449, 2005.

[18] "Egocentric Gesture Recognition Using 3D Convolutional Neural Networks for the Spatiotemporal Adaptation of Collaborative Robots," *Frontiers in Neurorobotics*, vol. 15, no. November, 2021.

[19] B. E. Olivas-Padilla, S. Manitsaris, D. Menychtas, and A. Glushkova, "Stochastic-biomechanic modeling and recognition of human movement primitives, in industry, using wearables," *Sensors*, vol. 21, no. 7, 2021.

[20] B. E. Olivas-Padilla, A. Glushkova, and S. Manitsaris, "Motion Capture Benchmark of Real Industrial Tasks and Traditional Crafts for Human Movement Analysis," *IEEE Access*, vol. 11, pp. 40 075–40 092, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10107408/