A Framework for a Closed Loop Control System of a Human Operator in a Manual Workstation

Maurizio Faccio*, Mattia Guidolin*, Leonardo Maretto*, Monica Reggiani*

** Dipartimento di Tecnica e Gestione dei Sistemi Industriali, University of Padova, Stradella San Nicola, 3 36100 – Vicenza – Italy (leonardo.maretto@phd.unipd.it, maurizio.faccio@unipd.it, mattia.guidolin@unipd.it, monica.reggiani@unipd.it)*

Abstract: Vision systems are being increasingly used in the digitalization process of industrial and manufacturing plants. Among them, marker-less Motion Capture (MOCAP) technology represent a valuable tool since it frees human operators from uncomfortable equipment on their body, allowing them to perform their activities as normal. We are currently implementing this technology at the Industrial Plants and Logistics Laboratory of the University of Padua, by applying a set of Intel Realsense depth cameras to a manual workstation and by linking them with a skeleton tracking software. The aim is to create a closed control loop that can monitor the activities performed by a human operator, offering real-time feedback. In this work we present a framework that describes the functioning of the closed loop and we present an example of its laboratory implmentation

Keywords: MOCAP, vision systems, Industry 4.0, digitalization

I. INTRODUCTION

Industry 4.0 (I4.0) has changed the shape of many factories worldwide. Lasi et al. (2014) define Industry 4.0 as a combination of two main mechanisms: an application pull, which combines a set of social, economic and political changes (ranging from individualization on demand to flexibility), and a technological push. The latter can be summarized in three main approaches: increasing mechanization and automation, digitalization and networking and miniaturization. The technological aspect, in particular, has acquired paramount importance within the Fourth Industrial Revolution. Many lists of I4.0 technologies are present in the literature. For example, Russmann et al. (2015) indicate nine enabling technologies: 1) Autonomous Robots, 2) Simulation, 3) Horizontal and Vertical Systems Integration, 4) Industrial Internet of Things, 5) Cybersecurity, 6) Cloud Computing, 7) Additive Manufacturing, 8) Augmented Reality and 9) Big Data and Analytics. Other more complex and structured lists exist, such as the one proposed by Frank et al. (2019), which indicates up to 33 different technologies, dividing them into four main groups: Smart Manufacturing Technologies, Smart Working Technologies, Smart Product Technologies and Smart Supply Chain Technologies. However, among the many existing lists, vision systems are not often cited despite their widening adoption and their promising results, as stated by Cohen et al. (2019). These systems are used both in human and robotized production systems. One of the most recently developed type of vision systems is Motion Capture (MOCAP). MOCAP is being used to digitalize some of the most difficult processes to be digitalized within a factory: manual processes. This fits perfectly within the digitalization and networking approach mentioned by Lasi et al. (2014). The increasing automation and the wide availability of technological instruments might call for an elimination of manual processes. However, in certain activities, the dexterity and cognitive abilities of a human operator are still unmatched by robots or automatic systems. For this reason, Industry 4.0 technologies, and MOCAP in particular, should be introduced alongside human operators, as collaborative elements.

In this paper we present a novel framework for a closed loop control system of a human operator. The aim of the control loop is to offer constant real-time feedback on the task sequence of an assembly procedure that is being performed by the operator. The online feedback system is well suited for an application of the implemented framework to support and control a human worker: this is the research area where the complete system is supposed to be applied. Support is meant as guidance before a certain task or activity while control is intended as monitoring after a task.

The rest of this work is organized as follows: Section II presents a literary review of the main contributions on the topic, Section III introduces the framework and describes each one of its steps, Section IV presents an example of a partial laboratory implementation of the framework, Section V defines the next steps of the research work and Section VI offers the final conclusions.

II. LITERARY REVIEW

According to Moeslund and Granum (2001), human motion capture can be defined as the process of capturing the large-scale body movements of a subject at some resolution. Human motion capture has three main areas of application: surveillance, control and analysis. The surveillance area covers applications where one or more subjects are being tracked over time and possibly monitored for special actions; the control area relates to applications where the captured motion is used to provide controlling functionalities; the third application area is concerned with the detailed analysis of the captured motion data.

Menolotto et al (2020) conducted a systematic literary review of the industrial applications of motion capture techniques. MOCAP industrial techniques were divided in two main groups: IMUs (Inertial Measurement Unit) and camera-based systems. The latter group was further divided in two subgroups: marker-based and marker-less. The study also showed that, among the 59 analysed contributions, only 6 of them focused on workers' productivity improvement.

Literature provides a variety of examples of camera-based marker-less motion capture systems in the industrial sector: Geiselhart et al (2016) adopted a multi-depth camera approach, based on Kinect v2 cameras, the Time-of-Flight principle and a skeletal tracking module. The flexibility and scalability of the proposed solution allows it to be used for production planning purposes; Otto et al. (2016) offer an example of application of marker-less motion capture in the automotive sector: multiple depth cameras are combined with different virtual and augmented reality tools to assess the progression of manual assembly tasks. The different combinations of systems are applied to different areas, such as accessibility and ergonomics assessment, interactive station layout planning and verification of assembly routines; Rodriguez et al. (2015) developed another combination of motion capture and mixed reality, although on a smaller scale: the operator is guided in the assembly tasks by a light projector, while only the hand movements on the workbench are tracked.

As suggested by the previous examples, the feedback to the operator is usually provided through augmented or virtual reality solutions. This is also shown in Dalle Mura et al. (2016), Sand et al. (2016) or Faccio et al. (2019).

A more limited amount of works explores the adoption of a combination of motion tracking and a feedback loop to the training of manual assembly tasks in the industrial sector.

Among these contributions, Muller et al (2016) developed SAW (Smart-Assembly-Workplace), which is designed to address the problem of knowledge transfer of assembly procedures between companies or subsidiaries located in different countries. The system is specifically designed to transfer knowledge on the assembly sequence of a bicycle e-hub to workers with very limited experience. The SAW adopts a Microsoft Kinect camera located on the top of the workstation to track the operator's hands. The SAW is also provided with an LCD screen which is meant to show work instructions. An expert worker defines the correct task sequence, picking sequence and the locations of the assembly tasks on the workbench. The expert's work sequence is also measured through the MTM (Methods-Time Measurement) theory, defining the ideal cycle time. Then, the unskilled worker will learn the assembly procedure at the SAW: instructions will be provided on the LCD screen, the hand movements will be tracked by the Kinect and compared with the movements of the expert operator, providing an error message any time a deviation appears. The unskilled operator will also be evaluated in terms of cycle time, comparing it with the standard time set by the expert. Despite resulting in longer cycle times compared to a traditional paper-instructions-based assembly process, the live-feedback with respect to the correct assembly sequence was deemed as a key benefit of SAW.

Kubo et al. (2019) focused on high-skill manual operations: a training system for metallic painting was developed. Knowledge transfer in metallic painting is particularly complex and the learning curves for beginners are particularly shallow. In this case, marker-based motion tracking was applied, with the aid of 34 sensors placed on the body of two expert operators and one beginner. At first, an analysis of the difference between the poses of the expert and the beginner was conducted. The output of the analysis was used to define a novel knowledge transfer system: the movements of an unskilled worker are compared in real time with the movements of an expert painter, and live feedback is provided in case of error.

Hodaie et al (2018) introduced a framework for the development of an intelligent tutoring system dedicated to manual activities. Manual work in factories is characterized by two main aspects: 1) manipulation of physical objects, performed either by hands or with tools, 2) a procedure than needs to be followed, either reading instructions, shadowing an expert operator or by simple knowledge of the task sequence. The authors define a manual-procedural activity as a sequence of multiple steps that must be performed in a specific order, with each step involving manipulation of physical objects. Therefore, learning a manual procedural activity involves acquiring both procedural knowledge about the steps as well as motor skills for manipulating objects. A system that acts as a trainer for manual procedural activities should hence provide guidance through the steps and offer personalized feedback in case of error. The proposed solution has two modules: an expert module, that is used to track the movements of an expert operator, and a trainee module. In the latter, an unskilled worker is guided through the procedure by instructions projected on the workbench and is offered constant feedback based on motion tracking of the hand movements performed by a depth camera.

In this context, we propose a novel framework for a control loop of a manual operator. Camera-based marker-less motion capture is at the core of the framework since it is used to track the operator's movements: it is implemented through a combination of depth cameras and a skeleton tracking software. The system follows the Control Volumes (CVs) approach, introduced by Faccio et al. (2019). This approach gives flexibility to the framework, since the definition of the CVs can be decided *a priori*, by an expert operator or even by an unskilled worker. Finally, the real-time feedback allows the framework to be applied for training and support and control purposes.

In terms of the proposed novelty, this work falls within an understudied area of application of MOCAP techniques, as mentioned by Menolotto et al (2020). Within this area, the novelty of this paper lies both in the approach, through the definition of a real-time control loop based on control volumes, as well as its application to training and support and control of a human operator in a manual workstation.

III. The framework

The proposed framework is shown in Figure 3, shown in Appendix A. This framework provides the main steps towards the definition of a control loop for manual assembly processes performed by human operators. The loop is implemented through a combination of software and hardware and is based on the introduction of a set of Control Volumes: the operator's position, the location of the control volumes and the task sequence are compared and, according to the results, a real time feedback is offered to the operator.

At the centre of the control loop there is a human operator who is performing an assembly process on a manual workstation. The assembly process that most ideally would fit with the proposed control volume approach is a process with a high frequency of picking activities or where different tasks need to be performed in different areas of the workbench. The adoption of vision systems eliminates the need to wear uncomfortable equipment, allowing the operator to move freely without restrictions.

A set of depth cameras are used to record the operator's movements within the workspace. According to the specific characteristics of the work cycle, a single camera or a multi-camera layout is adopted. A single camera can work if the performed activities are limited to the workbench and there is very little obstruction: it may work for assembly operations where the focus is only on the hands, arms or upper body. A multicamera setup, on the other hand, works very well in case of physical obstructions and can enlarge the monitored area: depending on the single depth camera performance, a working area of at least 2×2 m can be monitored. In this way, it is possible to track not only simple assembly tasks that are performed on the workbench but also picking activities from storage locations which are placed behind the operator's shoulders or outside the workbench. However, a multi-camera setup brings with it more complex calibration, synchronization and fusion procedures.

Based on the stream of frames provided by the depth cameras, a skeleton tracking software is applied to capture the operator's movements in the working area. The operator's body is discretized in a set of keypoints that represent important body parts or joints. This allows to estimate in real time the pose of the operator: in particular, the output of the application of the skeleton tracking software are the coordinates (x,y) of the keypoints which represent the operator's pose. Generally, many skeleton-tracking software can also give 3D pose information. As a matter of fact, this data is reconstructed according to a combination of the pose and the values of the other two coordinates, x and y. This estimation, however, lacks precision: therefore, in the framework coordinate z is measured by the depth cameras.

Depth is measured by the cameras for the entire frame. What matters in this case, however, is the depth estimation of the keypoints of the model for the operator's body. As a consequence, an additional step at software level is required: the keypoints with their (x,y) coordinates are superimposed over the frames with the depth estimation; each keypoint is then associated with the respective depth value. In this way, each keypoint is finally associated with a three-dimensional set of coordinates, (x,y,z).

Before introducing the control volumes, that are at the centre of this framework, it is necessary to define the other input of the control loop: the task sequence. The manual assembly operation should be broken down into a series of simple tasks, going as deep as defining the picking of a certain item from its own container. Then, each task is assigned to a certain position on the workbench and around that position a control volume is defined: for a picking-intensive activity, the control volumes should be located around the containers or boxes where the items are picked from; moreover, it is possible to divide the workbench in different areas and assign different tasks to each area: in this case the Control Volumes will be located around each one of the defined areas. The Control Volumes are simply defined as a set of coordinates and can be shaped in different ways: spheres, cubes, cuboids etc. The definition of the Control Volumes is performed at software level, in order to store their coordinates and to quickly transfer them to the tracking algorithm.

Once both the coordinates of the operator and the control volumes are available, it is possible to compare them in real time. At a given point in time, the position of a certain set of keypoints of the operator are tracked and compared to the coordinates of the Control Volume that is associated with the task that is supposed to be performed according to the sequence of the specific activity. Two outcomes are possible:

1) the keypoints are located within the right control volume: in this case, a confirmation message is sent to the operator, reinforcing the correctness of the task sequence that is being followed. Moreover, once the task is finished, it is possible to provide the operator with a suggestion on the subsequent task, indicating the next control volume that needs to be addressed

2) the keypoints are located within the wrong control volume: an error message is sent to the

operator, possibly with the addition of a reminder of the correct sequence to follow.

The real-time feedback can be delivered to the operator in different ways, according to the available technical devices, but always keeping the basic requirement of offering error and confirmation messages according to the relative position of the operator and the control volumes. We provide here two examples of possible solutions:

a) A combination of smart wearable wristband and workstation-mounted projector: the projector indicates the next correct task by projecting the correct control volume on the workstation while the wristband communicates the successful completion of the task to the operator through its vibration pattern.
b) A workstation-mounted screen: the screen can show to the operator the control volumes and the correct task sequence, as well as prompting error and confirmation messages thus offering the required real-time feedback.

The feedback influences the operator, who adjusts his behavior according to the received suggestions: the movements are again tracked by the depth cameras and the skeleton tracking software, closing the control loop.

Finally, since the cameras and the software can continuously track the operator's movements and keypoints and can record the time instant in which a certain task is performed, the system is able to extract a set of relevant outputs: number of errors performed during the entire assembly and for each specific task, task duration, spaghetti charts of certain body parts of the operator etc. This allows to perform in depth data analysis of the assembly process.

IV. LABORATORY IMPLEMENTATION EXAMPLE

The framework is general and allows different hardware and software combinations to be applied. The control loop is currently being implemented at the Industrial Plants Laboratory of the Department of Management and Engineering at the University of Padua. In this section, an example of the adopted software and hardware technologies is presented.

I. Depth Cameras

The adopted depth cameras are Intel RealSense D435i, shown in Figure 1.

Fig. 1. Intel RealSense D435i depth camera (Intel® RealSense™ Depth and Tracking Cameras., 2022)

These cameras adopt stereo vision as a means to measure depth. Two imagers are the starting point for the stereo vision implementation: they are two identical camera sensors, one located on the left of the camera and the other on the right, with 1920 × 1080 active pixels and configured with the same settings. Both imagers capture the same frame. The frame data are sent to the vision processor of the camera (Intel RealSense Vision Processor D4), which is assigned the task of estimating the depth values: the same points in the left and right image are correlated and then a triangulation approach is followed, calculating the depth values depending on the shift in pixels between those same points since the distance between the left and right imager is known. The single depth pixel values are then elaborated and converted into a depth frame. The camera has an operating range of 0.3m to 10m and it guarantees a depth accuracy of 98% for distances lower than 2m.

At the moment the setup includes just one camera but the D435i is already set for multicamera use since it is provided with a built-in synchronization mechanism.

II. Skeleton Tracking Software

The adopted body pose estimation software is OpenPose, which is an open-source real time system for multi-person 2D pose detection, developed by Cao et al. (2019). OpenPose falls within the category of bottom-up approaches for pose estimation software: body parts are recognized first as a set of keypoints, which are later assembled as limbs according to different association techniques (Li et al., 2019). This approach works well in the case of multiple subjects in the frame or in presence of obstructions. Since many possible obstructions are present in a manual workstation, OpenPose represents a viable option for the skeleton tracking software choice within the framework.

The input of the software is a color image: it can be either a single picture or a frame of a video. The image is then fed to a Convolutional Neural Network (CNN) that adopts 3 consecutive 3×3 kernels. The CNN jointly predicts a set 2D of confidence maps for the detection of the single body parts and another set of 2D vector fields, called Part Affinity Fields (PAFs), which are used to associate the detected body parts. The confidence maps and the PAFs are then matched according to a greedy inference process, resulting in the full body poses: in this way, the 2D keypoints of the figures in the images are finally computed.

The 2D confidence maps are a tool that is used to represent the probable position of each one of the keypoints that are used to discretize the body of the persons in the frame: for each body part, each pixel of the frame is associated with the probability of having that body part located in that pixel. Given the recognized body parts, PAFs are used to assemble the full body poses by computing a confidence measure of the association of each pair of keypoints: for each pixel located in the area belonging to a particular limb, a 2D vector encodes the direction that points from one part of the limb to the other. Each type of limb has a corresponding PAF joining its two associated body parts.

Different versions of the body pose estimation model exist: MPI, COCO, BODY_25 etc. BODY_25 is the one adopted in our proposed implementation: the body of the operator is discretized in 25 keypoints, as shown in Figure 2. In the proposed control-volume approach, the candidate keypoints to track the task sequence completion are number 4 and 7, which correspond respectively to the right and left wrist.

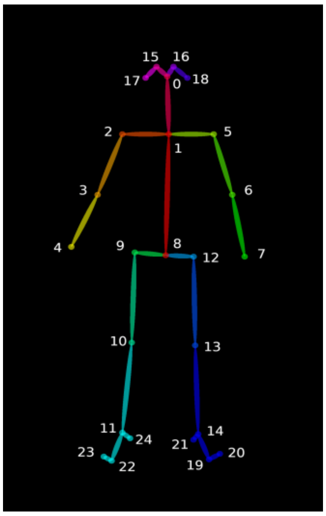

Fig. 2. BODY_25: the 25 keypoints are numbered from 0 to 24 (GitHub., 2022)

III. ROS

An internally developed algorithm is exploited to integrate the (*x*,*y*) coordinates estimated by OpenPose with the depth information of the RealSense camera, allowing to estimate the 3D position of each body keypoint. To ensure real-time performance and enable simultaneous use of multiple sensors, the software is based on the Robot Operating System (ROS) (Quigley et al. 2009). ROS is an open-source meta-operating system that includes a large variety of libraries and tools for developing software modules that communicate with each other in a loosely coupled, multiprocess, distributed environment. A ROS-based system is composed of nodes, which are processes that perform a computation from running algorithms for interfacing with sensors or actuating devices. Nodes communicate with each other using a publish/subscribe model (i.e., a node that produces data publishes them on a named topic). ROS takes charge of distributing the data only to the nodes that previously subscribed to this topic.

V. Research Roadmap

As mentioned in the previous section, the implementation process of the framework is still undergoing. Therefore, the definition of a research roadmap that clearly states the future implementation steps and the successive testing campaigns is required.

First of all, the setup of the systems needs to be completed, defining all the remaining hardware and software implementations.

Within the setup definition, the first issue that needs to be addressed is the decision on the type of software to be used for the comparison of the coordinates between the operator and the control volumes. The chosen software should guarantee high computational performance in order not to reduce the frame rate of OpenPose, thus maintaining the close-to-real-time control of the operator's position.

With regards to the feedback system for the operator, many options are available, as mentioned in Section III. The proposed idea for the implementation is to adopt a simple webcam that captures the working area, with a specific focus on the zone where the operator's hands are located. This webcam will produce a live stream that will be sent to a screen positioned on the side of the workstation. The images from the webcam will be enriched with a graphical representation of the control volumes. The screen will also display the error and confirmation messages related to the correct task sequence. In such manner, the operator will be provided with constant visual feedback.

The control volumes definition is the last step of the implementation of the framework and needs to be tuned according to the specific characteristics of the experiment. As a matter of fact, the type of object to assemble, the number of its components, the lot size etc. are all factors that influence the dimensions of the control volumes first and consequently their location. Therefore, these parameters need to be defined before the creation of the control volumes.

Once the setup is ready and the experimental parameters are clear, the campaign can start. The idea is to study how such a control loop can influence the learning path of a manual operator. In particular, the focus will be on the comparison between an operator who performs the assigned assembly tasks with the aid of paper instructions and an operator who is subject to the feedback of the proposed control loop. They will be analyzed in terms of performance, especially through their respective learning curves. The design of the experiments will also influence the set of output parameters to be extracted from the loop.

VI. Conclusions

In this paper we presented an original framework for the definition of a control loop of a manual operator. A workstation is digitalized with the aid of one or more depth cameras which, in combination with a skeleton tracking software, provide the spatial coordinates of the human operator. Those coordinates are then compared with the positions of a set of pre-defined Control Volumes, which are used to track the task sequence progression. Real-time feedback is provided to the operator: an error message is sent if the operator is working on the wrong Control Volumes while a confirmation message and the suggestion of the next step are prompted if the task sequence is correct. The implementation of a system that follows the guideline of the framework is currently underway at the Industrial Plants Laboratory of the University of Padua. According to the research roadmap, the complete system will be applied to study the learning curves of manual operators in presence of a real-time feedback.

REFERENCES

[1] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-. & Sheikh, Y. 2021, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172-186.

[2] Cohen, Y., Faccio, M., Pilati, F. & Yao, X. 2019, "Design and management of digital manufacturing and assembly systems in the Industry 4.0 era", *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 9, pp. 3565-3577.

[3] Faccio, M., Ferrari, E., Galizia, F.G., Gamberi, M. & Pilati, F. 2019, "Real-time assistance to manual assembly through depth camera and visual feedback", *Procedia CIRP*, pp. 1254.

[4] Frank, A.G., Dalenogare, L.S. & Ayala, N.F. 2019, "Industry 4.0 technologies: Implementation patterns in manufacturing companies", *International Journal of Production Economics*, vol. 210, pp. 15-26.

[5] Geiselhart, F., Otto, M. & Rukzio, E. 2016, "On the Use of Multi-Depth-Camera Based Motion Tracking Systems in Production Planning Environments", *Procedia CIRP*, pp. 759.

[6] GitHub, 2022. OpenPose. [online] Available at: <https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html> [Accessed April 2022].

[7] Hodaie, Z., Haladjian, J. & Bruegge, B. 2018, *TUMA: Towards an intelligent tutoring system for manual-procedural activities*.

[8] Intel® RealSense™ Depth and Tracking Cameras. 2022. *Depth Camera D435i*. [online] Available at: <https://www.intelrealsense.com/depth-camera-d435i/> [Accessed April 2022]

[9] Kubo, M., Hida, T., Nakajima, R. & Matsumoto, T. 2019, "Visualization of skilled worker motion and its effect on product quality and the design of skill training systems in metallic painting operations", *Industrial Engineering and Management Systems*, vol. 18, no. 1, pp. 25-34.

[10] Lasi, H., Fettke, P., Kemper, H.-., Feld, T. & Hoffmann, M. 2014, "Industry 4.0", *Business and Information Systems Engineering*, vol. 6, no. 4, pp. 239-242.

[11] Li, M., Zhou, Z. & Liu, X. 2019, "Multi-Person Pose Estimation Using Bounding Box Constraint and LSTM", *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2653-2663.

[12] Menolotto, M., Komaris, D.-., Tedesco, S., O'flynn, B. & Walsh, M. 2020, "Motion capture technology in industrial applications: A systematic review", *Sensors (Switzerland)*, vol. 20, no. 19, pp. 1-25.

[13] Moeslund, T.B. & Granum, E. 2001, "A survey of computer vision-based human motion capture", *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231-268.

[14] Müller, B.C., Nguyen, T.D., Dang, Q.-., Duc, B.M., Seliger, G., Krüger, J. & Kohl, H. 2016, "Motion Tracking Applied in Assembly for Worker Training in different Locations", *Procedia CIRP*, pp. 460.

[15] Mura, M.D., Dini, G. & Failli, F. 2016, "An Integrated Environment Based on Augmented Reality and Sensing Device for Manual Assembly Workstations", *Procedia CIRP*, pp. 340.

[16] Otto, M., Prieur, M., Agethen, P. & Rukzio, E. 2016, "Dual Reality for Production Verification Workshops: A Comprehensive Set of Virtual Methods", *Procedia CIRP*, pp. 38.

[17] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., … & Ng, A. Y. (2009, May). ROS: an open-source Robot Operating System. In ICRA workshop on open source software (Vol. 3, No. 3.2, p. 5).

[18] Rodriguez, L., Quint, F., Gorecky, D., Romero, D. & Siller, H.R. 2015, "Developing a Mixed Reality Assistance System Based on Projection Mapping Technology for Manual Operations at Assembly Workstations", *Procedia Computer Science*, pp. 327.

[19] Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. & Harnisch, M. 2015, "Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries", *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries.*

[20] Sand, O., Büttner, S., Paelke, V. & Röcker, C. 2016, *smARt.Assembly – Projection-based augmented reality for supporting assembly workers*.

Appendix A. FRAMEWORK REPRESENTATION

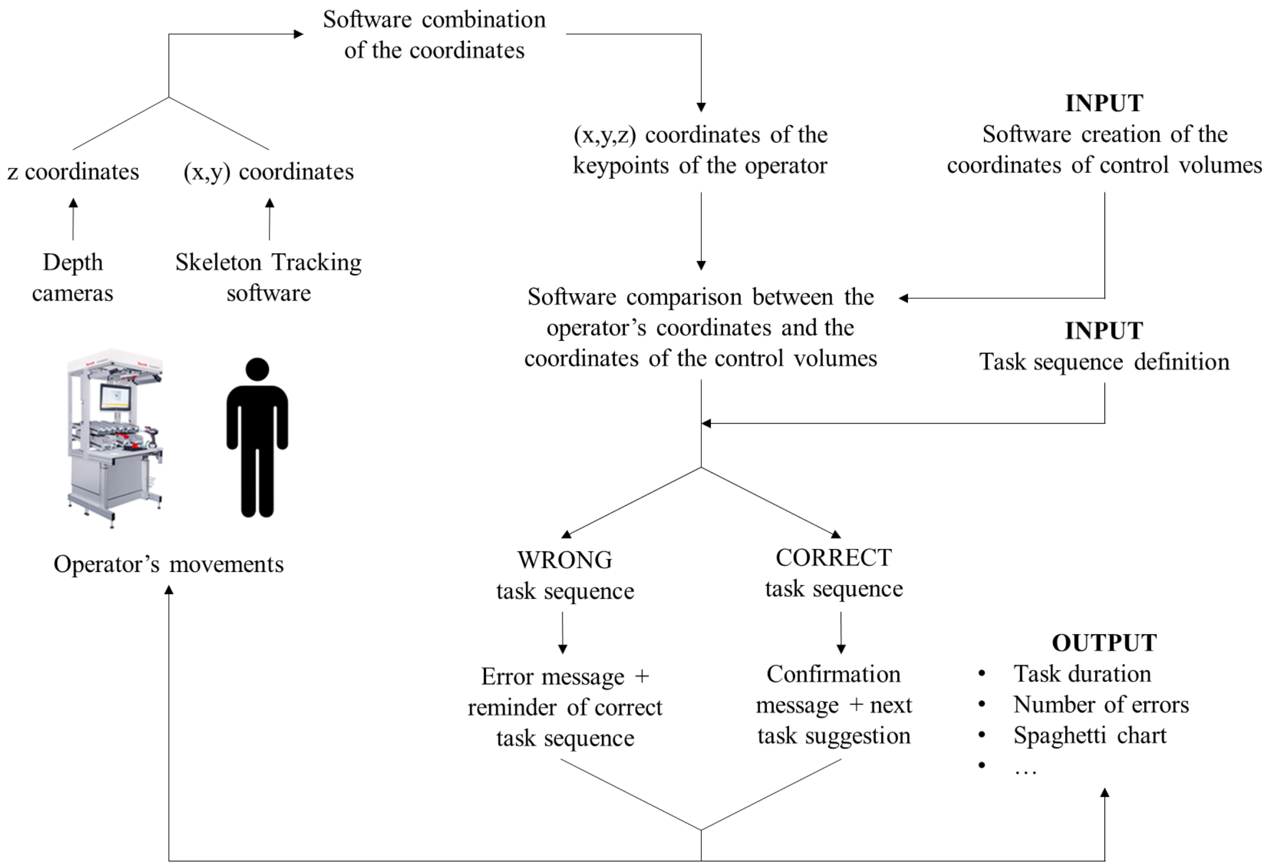

Fig. 3. Framework for a closed loop of a human operator in a manual workstation