Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 81 (2019) 1254–1259

www.elsevier.com/locate/procedia

52nd CIRP Conference on Manufacturing Systems

# Real-time assistance to manual assembly through depth camera and visual feedback

Maurizio Faccio[a], Emilio Ferrari[b], Francesco G. Galizia[a], Mauro Gamberi[b], Francesco Pilati[b,*]

[a] *University of Padua, Department of Management and Engineering, Stradella, San Nicola 3, 36100 Vicenza, Italy*
[b] *University of Bologna, Department of Industrial Engineering, Viale del Risorgimento 2, 40136 Bologna, Italy*

* Corresponding author *E-mail address:* francesco.pilati3@unibo.it

**Abstract**

The current fourth industrial revolution significantly impacts on production processes. The personalized production paradigm enables customers to order unique products. The operators assemble an enormous component variety adapting their process from product to product with limited learning opportunities. Digital technologies are increasingly adopted in production processes to improve performance and quality. Considering this framework, this research proposes a hardware/software architecture to assist in real-time operators involved in manual assembly processes. A depth camera captures human motions in relation with the workstation environment whereas a visual feedback guides the operator through consecutive assembly tasks. An industrial case study validates the architecture.



*Keywords:* aided assembly; real-time; depth camera; motion capture; assisted picking; error detection; feedback; industry 4.0; smart factory; digital manufacturing

## 1. Introduction and literature review

In the current era characterized by the advent of Industry 4.0, the success of an industrial company is strongly linked to its capacity to quickly adapt to the emerging social, technological and economic conditions [1, 2]. High flexibility, increasing customization, flexible batches and short product life cycles are driving the transition from the traditional production systems to the so-called Next Generation Production Systems (NGPSs) [3, 4]. In these systems and according to the Industry 4.0 strategy, the human operators play a crucial role since they represent the most flexible resource in assembly and manufacturing systems being adaptive and responsive to the increasingly challenging work environment [5]. The requirements that the operators have to face with on the shop floor are the increased product complexity, the shortened product development cycles and the presence of a huge number of product variants. These factors inevitably affect the performance of manual operations and make the manufacturing and assembly processes more susceptible to human errors, resulting in delays, defects and/or poor product quality [6]. In this context, the implementation of adequate qualification measures at the organizational and technological levels is required to enable the human operators to efficiently perform their tasks minimizing errors and product defects, during manufacturing and assembly processes. The existing assistance systems for parts picking during assembly operations can be classified in Pick-by-Light and Pick-by-Vision systems. In Pick-by-Light systems, a light is attached to every component storage location in a rack. The light supports the human operators showing which and how many part to pick [7, 8]. However, Pick-by-Light solutions need to be physically


10.1016/j.procir.2019.03.303

installed on the assembly workstation, making the system inflexible and static. Moreover, while these systems provide an effective spatial feedback to the user, they are too expensive and unsuitable in small installation with high turnover [9]. On the other hand, Pick-by-Vision systems are based on the application of Human-Machine Interface (HMI) and Augmented Reality (AR) techniques. AR deals with the enrichment of the real world by virtual overlays and its main ability is to show virtual instructions directly in the operator's field of view [10]. Such systems typically use smart glasses to display the component to pick. One advantage is that no installation on racks is required and a broader picking area can be monitored with one single device [7]. However, no feedback is provided to the operator concerning the correctness of performed action. Recent literature contributions are moving toward the development of marker-less devices, by using computer vision systems and depth cameras [11, 12], to accurately evaluate in real-time the operator interaction with the surrounding environment. During the last few years, relevant industrial companies as Boeing, Volkswagen and BMW started to conduct pilot studies to implement AR techniques in their production lines in order to improve their manufacturing and assembly processes [6, 13]. An interesting study is in Gorecky et al. [14]. They present a methodology for a computer-assisted workstation, supporting manual assembly of small parts by applying AR and advanced sensor technology. In particular, the system provides virtual instructions and guides the user step-by-step through the manual assembly process. Rodriguez et al. [6] develop a mixed reality, i.e. virtual and augmented, assistance system based on projection mapping technology to support the execution of manual operations at assembly workstations. Gorecky et al. [5] introduce the design, implementation and evaluation of an advanced virtual training system by using Microsoft Kinect$^{TM}$ as motion tracking system. Korn et al. [15] propose a context-aware assistive system for industrial workplaces able to capture in real-time the user's movements and provide adequate feedbacks. The system makes use of projectors to display the feedback within the corresponding area on the workspace. Sand et al. [16] present smART.assembly, a projection-based AR assistance system for industrial applications. Such system projects digital guidance information in terms of picking information and assembly data into the physical workspace of a user. The use of projection techniques eliminates the use of smart glasses that have some drawbacks such as limited field of view and low wearing comfort. Radkowski [17] introduces a 3D object tracking method for an AR assembly assistance application. The proposed method relies on the use of 3D feature descriptors and point cloud matching with the iterative closest points (ICP) algorithm. Wang et al. [18] propose a tracking method based on the combination of point cloud and visual feature in order to improve the applicability and robustness of three-dimensional tracking method of an AR assembly guiding system for mechanical products.

Finally, in the latest years, several researchers focused their effort on the development of software solutions, ICT systems and digital architectures to guide, ease and assist manual assembly processes. Bader and Aehnelt [19] propose an assistance system to track and guide operators during the execution of manual assembly tasks leveraging a technology widespread in several manufacturing companies, e.g. RFID, and processing the acquired data with a probabilistic model. Despite this research is very promising to support the workers during task execution, it does not track nor digitalize any human movements performed to execute those tasks. Otto et al. [19] propose a comprehensive set of virtual methods to support the manual production processes of the automotive industry. Between the multiple solutions proposed, a set of multiple depth cameras aim at full body tracking of human operators. However, limited attention is paid to the assembly of small components in static and confined workspaces, e.g. workbench. Dalle Mura et al. [21] overcome this limitation by focusing on manual assembly on workbench. These authors adopt a force sensor placed under the workbench to monitor the assembly process in terms of type and sequence of picked components by collecting force and torque data with respect to a shared reference system. An LCD and a head mounted display are alternatively employed to give the operator a real-time feedback about the correctness of the performed tasks in an AR environment. However, the mapping of the assembly area to evaluate the interaction between the operator and the products or components requires a non-negligible amount of time and competences to be set-up. Westerfield et al. [22] proposes a ready-to-use and plug-and-play intelligent tutoring system targeted to the assistance of operator training to perform novel manual assembly tasks through a customized AR feedback device, e.g. a head mounted display. The innovation proposed promises to significantly reduce the training phase duration but is not aimed to be used for day-by-day and traditional mounting processes.

Starting from this scenario, this paper proposes an original hardware/software architecture to assist and guide in real-time the human operators involved in manual assembly processes. A depth camera tracks human motions in the workstation environment and a visual feedback assists the human operator during the performance of consecutive assembly tasks. Compared to traditional electromagnetic-based and opto-electronic motion tracking systems, the considered depth camera sensor is a low-cost, portable and marker-less solution [23-25]. The reminder of this paper is organized as follows. Next Section 2 describes the proposed hardware/software architecture for human operator real-time tracking which results in an original app developed for assistance during assembly tasks presented in Section 3. An extensive experimental campaign is carried out to investigate the performances of the develop architecture and it is summarized in Section 4. The adoption of the developed hardware/software architecture to a case study representative of the assembly operations of a centrifugal electric pump is in Section 5. Section 6 concludes the paper with final remarks and future research opportunities.

## 2. Hardware and software architecture

This Section presents the technology adopted and the digital procedure developed to provide a unique hardware and software architecture able to assist in real-time the operators performing assembly activities with multiple possible variations of the executed tasks due to the wide variety of the final products manufactured.

Concerning the hardware architecture, the depth camera technology is adopted as a flexible, adaptable and noninvasive solution to monitor and track the operator movements in a 3D environment. The time-of-flight technology is selected levering a pair of RGB color and infrared depth sensors. Furthermore, the camera firmware has been customized to maximize the motion capture accuracy in an industrial environment affected by flashing lights and metallic materials. The most relevant features of the selected optical sensors are listed in the followings. The depth camera adopted in this research is not a Microsoft Kinect, but a commercial optical sensor recently developed and released which promises to offer distinctive performances.

- Depth device accuracy of 620x460 at 30 fps;
- RGB device accuracy of 1260x700 at 30 fps;
- Vertical/horizontal FOV: 60°/75°;
- Maximum/minimum tracking distance: 7.0m/1.3m;

The adopted depth camera enables to track the operator activities within most of the industrial workstations. The markerless technology does not require the worker to wear any cumbersome suit. The usage of optical sensors instead of wire connected devices frees the adoption of this hardware/software (HW/SW) architecture to almost every assembly station. Finally, the wide and deep camera field of view makes this motion capture solution adaptable to several workplace layouts overcoming possible tracking occlusions. The following Fig. 1 presents the front (left) and top views (right side) of the depth camera disposition for conventional (top) and non-traditional (bottom) workstation layouts. For these layout configurations one depth sensor is enough to adequately monitor the operation motions during task execution. The possibility to use just one camera is of major importance to avoid the fusion in real-time of data streams from multiple depth sensors, which would increase the complexity of the hardware/software architecture while reducing its reliability.

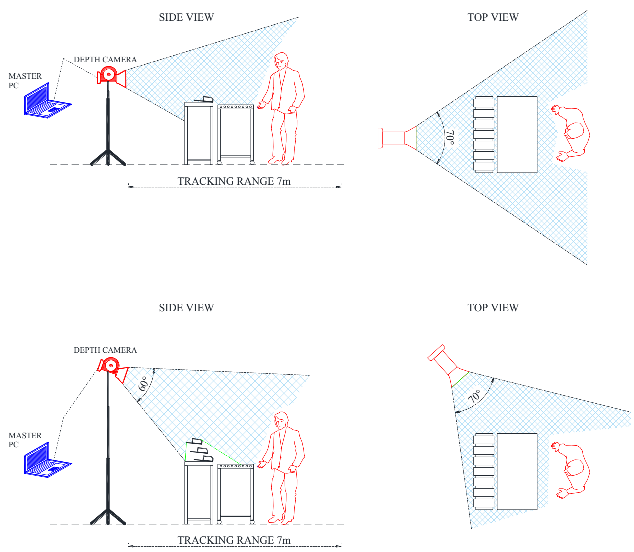


Fig. 1. Camera disposition for conventional (top) and non-traditional (bottom) layouts, front (left) and top views (right side).

The proper artificial intelligence algorithms encoded by the commercial depth camera manufacturer in the camera SDK are able to identify the human body profile within the sensor field of view compared to the scene background to evaluate the body joint 3D coordinates in a shared reference system. These algorithms evaluate at 30 Hz the absolute geometric coordinates (x,y,z) of each of the 20 joints detected by the camera with a theoretical precision of about a centimeter. However, the multiple tests performed in real industrial environments suggest that the actual accuracy of the body tracking solution provided by the commercial optical sensor manufacturer is approximately 4-5 cm.

The real-time location of the operator limbs is a valuable information if properly leveraged, in particular referring these measures to the workstation layout. The Control Volumes (CVs) concept is proposed for this purpose. A CV is a virtual object of any geometrical shape with defined dimensions and known 3D position within the monitored area. The CVs can be displaced on any relevant location of the productive environment, as on workbenches, inside boxes, around bins, on shelves, etc., ensuring a relevant flexibility. Checking if the 3D position of the operator body joints enters/exits to/from a particular CV enables to evaluate in real-time the interaction between the worker and his productive environment (Fig. 2). Finally, a visual feedback is provided to the worker through a monitor superimposing on the RGB live video the mannequin of the operator digitalization and its interaction with the virtual CVs validating right activities and highlighting wrong actions (see Section 4, Figs. 6-7).

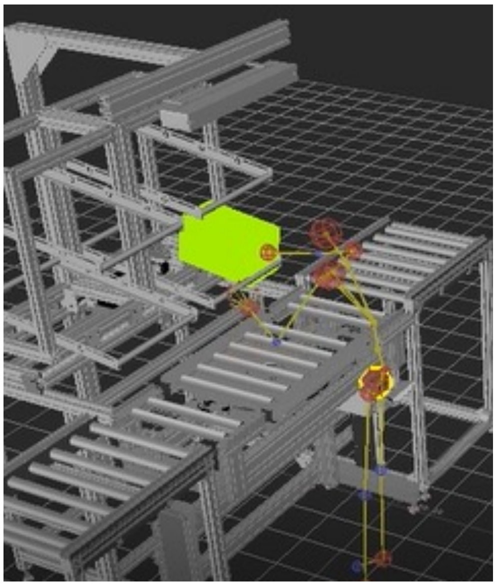

Fig. 2. Interaction between the operator hands and the CVs displaced on the workstation shelves.

## 3. Aided assembly app

The proposed HW/SW architecture is leveraged to develop an original digital application to aid the worker during manual assembly, in particular for low repetitive activities as sequentially mounting of multiple product varieties. The proposed app relies on three consecutive phases each distinguished by an execution frequency, a description and the output generated.

Phase 1: CVs positioning

*Execution frequency:* Once per workstation layout.

*Description:* The virtual CVs are displaced within the productive environment leveraging the motion capture of a human trainer which indicates the CV locations through proper gestures previously encoded in the app. For instance, in the component picking process every storage bin can be associated to a CV. The CV shape and dimensions are provided by an input file.

*Output:* Coordinates of the 3D position of every CV within the workstation.

Phase 2: Assembly training

*Execution frequency:* For every product variety (model or configuration) to be manufactured.

*Description:* The trainer executes the sequence of tasks required by the product variety to be assembled performing consecutive accesses with his hands to a subset of CVs. For instance, the app can assist a worker during the ordered picking of product components from several storage bins on the workbench. During this phase, the trainer performs the component picking from the different CVs following the sequence required by the considered product variety.

*Output:* Ordered lists of CVs to be accessed in sequence for every product variety.

Phase 3: Real-time assistance

*Execution frequency:* In real-time at 30 Hz.

*Description:* Assistance to the worker during assembly activities using as input the outputs of phases 2 and 3. Automatic identification of the next CV to access considering the product variety and the assembly tasks progression. Visual feedback to the operator through a monitor about the correctness of the executed action. For instance, the app can guide an operator through the sequential assembly tasks highlighting in the monitor the next CV to visit to pick the right component. The correct picking is validated through a positive feedback proposing through video the next CV to access. Incorrect picking actions are detected immediately preventing errors in the component mounting. The description of this phase is summarized by the flow diagram presented in the following Fig. 3.

*Output:* Assisted assembly through visual feedback preventing possible mounting errors.

## 4. Experimental campaign

An extensive experimental campaign is performed at the Digital Production Laboratory of the Department of Industrial Engineering – University of Bologna (Italy) to investigate the accuracy, reliability and usability of the developed HW/SW architecture and the related aided assembly app. 12 different operators (7 male, 5 female; height between 1.62 and 1.91 m) performed an identical set of 16 tasks (duration between 0.2 and 3.4 mins) to assemble 3 different products requiring 28 different components. The workstation layout is comparable to the one presented in Fig. 1, e.g. a workbench with component storage boxes (14.0 x 26.0 x 12.5h cm) in front the operator. Different layout configurations have been tested varying the level number of component boxes stacked on top of each other from 1 to 3 and the camera position. For this latter, the frontal distance between the camera and the operator has been varied between 1.5 and 5.0 m, the lateral one between 0 and 2.5 m and the camera high between 1.0 and 3.0 m. Furthermore, concerning the end user feedback, an anonymous questionnaire has been filled in by the involved operators. The questionnaire asked to evaluate using a Likert scale the easiness, tiredness and control perception in using the aided assembly app, as well to propose any possible improvement to the developed app. Finally, the

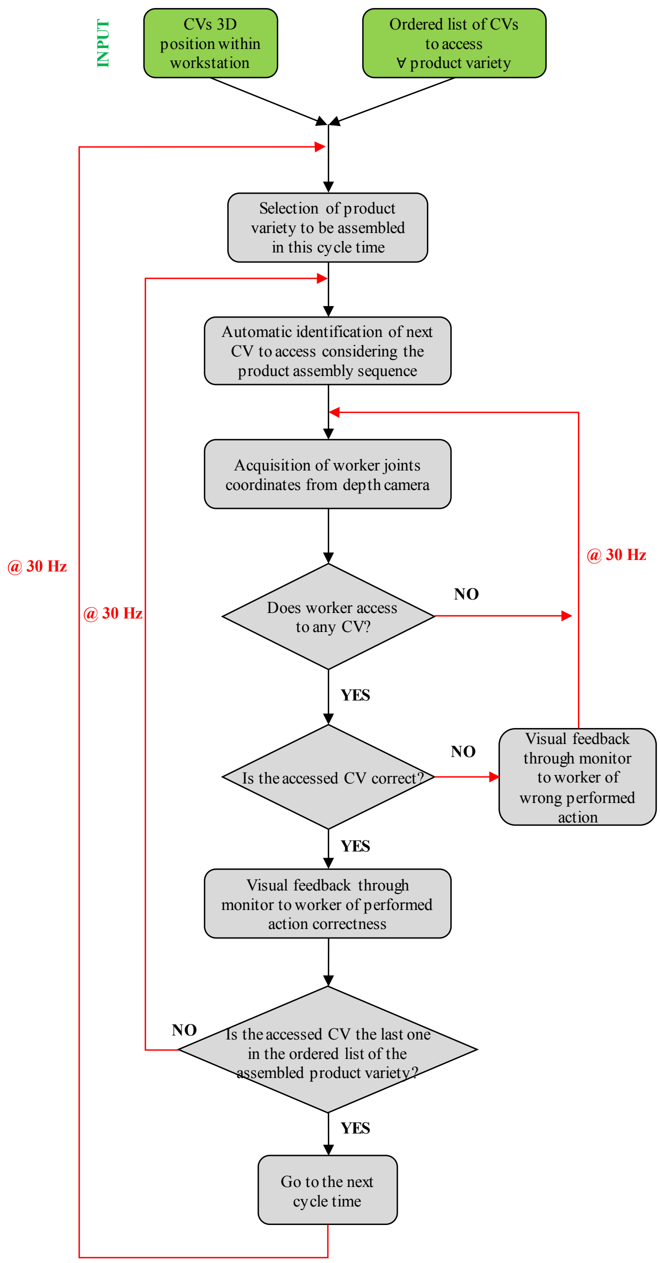


Fig. 3. Flow diagram of the procedure for the worker real-time assistance.

manual assembly performance improvement in using the developed HW/SW architecture is evaluated measuring the execution time, the number of wrong CV accessed and the distance travelled by the operator hand for the different tasks, with and without the assistance of the developed app. For sake of brevity, this manuscript proposes in the following just a summary of the experimental campaign results, however the authors would be more than glad to provide all the desired details, if contacted. The optimal HW/SW configuration is experienced with a camera exactly in front of the operator at 2.8 m distance and 1.6 m height which ensures to distinguish with 96% of accuracy (4.8 cm average error of body joint tracking) the picked components from the boxes even attached to each other and stacked in two levels. The operator performance improvement is represented by a reduction of the assembly time between 6% and 18% and of hand travelled distance between 3% and 16%. The operator feedback suggests how this HW/SW architecture significantly eases the performed tasks (5.8 out of 7 on Likert scale) with almost no impact on the fatigue (3.2/7) but suggesting to not neglect the perception of being controlled while performing their working activities (5.1/7).

## 5. Case study and results

### *5.1 Case study description*

To test and validate the proposed hardware/software architecture along with the aided assembly app, an industrial case study of a European manufacturer is presented, representative of a manual assembly process of a water pump. The pump is characterized by 6 components and the assembly process has to be performed within the cycle time of 4.8 min/pcs. Fig. 4 shows the product and its main components.

Fig. 4. Water pump assembly and main components.

In the considered case study, one male operator 173 cm high performs the 16 tasks required for the assembly operation. In particular, the product is assembled over a 1.4 x 1.2 x 0.9 (h) meters workbench while the components are positioned over the workbench in 18 x 26 x 12 (h) cm blue boxes according to their type. In the developed aided assembly app, each blue box is modelled as a control volume to monitor within the assembly process. The adopted depth camera is positioned in front of the operator working area, together with a display that shows in real-time the camera recording and tracks the operator within the duration of the whole process without any interruptions. Such camera is able to capture an area of 3.2 x 2.4 meters.
The proposed app performs a digitalization of the operator human body and shows on the screen its skeleton characterized by some relevant joints, white points in Fig. 5.

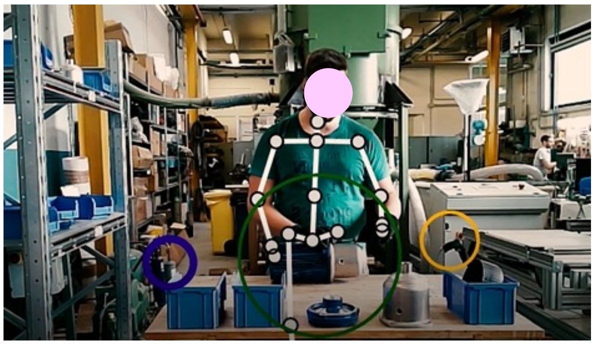

Fig. 5. Main elements of the aided assembly app.

Fig. 5 shows the main relevant elements provided by applying the developed app. In particular, the operator working area is highlighted by a green circle. Moreover, having in input the work cycle of the product to assemble as well as the component type that each blue box holds, the methodology supports the operator in the identification and visualization of the next control volume to access, identifying it with a yellow circle.

### *5.2 Main results*

The aided assembly app supports in real-time the operators during the execution of manual tasks providing a positive feedback in case of access to the right control volume. In this case, a low green circle appears near the right control volume together with the yellow word “OK” that appears on the screen (Fig. 6).


Fig. 6. Positive feedback to the human operator.

The app provides support within the assembly process even in the case of wrong access to a control volume. Consider the example in Fig. 7. The operator has to access to the control volume characterized by the yellow circle since it contains the component type required to perform the next task of the water pump assembly process. However, the operator puts his hand in another blue box. In this case, a red circle appears near the wrong control volume to which the operator accessed together with the red word “NO” that appears on the screen.

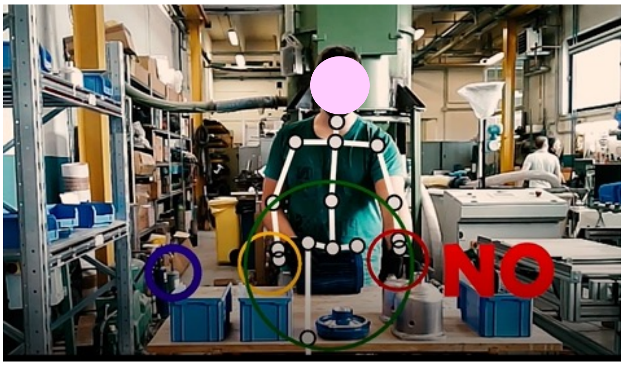


Fig. 7. Negative feedback to the human operator.

The yellow circle near the correct CV is shown on the screen as long as the operator performs the right access. Providing in real-time positive and negative feedbacks to the operators performing manual assembly tasks, the proposed methodology allows a reduction of errors and delays and a greater ability to manage an ever-increasing number of product variants and product customizations.

## 6. Conclusions

The transition from mass customization to personalized production paradigm determines a several challenges for industrial operators. High flexibility, dynamic market demand, flexible batches and short product life cycles are among the key factors driving the transition from the traditional manufacturing systems to the so-called Next Generation Production Systems (NGPSs). In this emerging context characterized by the advent of Industry 4.0, customers ask for an increasing number of variants and product customization. These factors affect the performance of manual operations and make the manufacturing and assembly processes more susceptible to human errors and delays. Starting from this scenario, this paper proposes an integrated HW/SW architecture to assist and guide in real-time through an original app the operators involved in manual assembly tasks. A depth camera captures human motions and a visual feedback guides the operator through consecutive assembly tasks, providing both positive and negative feedbacks. An industrial case study representative of an assembly process of a water pump validates the proposed architecture and app.

## References


[1] Bortolini M, Faccio M, Gamberi M, Pilati F. Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes. Computers & Industrial Engineering 2018; [in press, DOI: 10.1016/j.cie.2018.10.046].

[2] Faccio M, Ferrari E, Gamberi M, Pilati F. Human Factor Analyser for work measurement of manual manufacturing and assembly processes. The International Journal of Advanced Manufacturing Technology 2019; [in press, DOI: 10.1007/s00170-019-03570-z].

[3] Molina A, Rodriguez CA, Ahuett H, Cortés JA, Ramirez M, Jimenez G, Martinez S. Next-generation manufacturing systems: key research issues in developing and integrating reconfigurable and intelligent machines. International Journal of Computer Integrated Manufacturing 2005;18(7):525-536.

[4] Bortolini M, Galizia FG, Mora C. Reconfigurable manufacturing systems: literature review and research trend. Journal of Manufacturing Systems 2018;49:93-106.

[5] Gorecky D, Khamis M, Mura K. Introduction and establishment of virtual training in the factory of the future. International Journal of Computr Integrated Manufacturing 2017;30(1):182-190.

[6] Rodriguez L, Quint F, Gorecky D, Romero D, Siller HR. Developing a mixed reality assistance system based on projection mapping technology for manual operations at assembly workstations. Procedia Computer Science 2015;75:327-333.

[7] Gunthner WA, Blomeyer N, Reif R, Schedlbauer M. Pick-by-Vision: augmented reality unterstutzte kommissionierung.

[8] Fasth-Berglund A, Stahre J. Cognitive automation strategy for reconfigurable and sustainable assembly systems. Assembly Automation 2013;33(3):294-303.

[9] Baumann H. Order picking supported by mobile computing. Ph.D Thesis, University of Bremen.

[10] Vignais N, Miezal M, Bleser G, Mura K, Gorecky D, Marin F. Innovative system for real-time ergonomic feedback in industrial manufacturing. Applied Eergonomics 2013;44(4):566-574.

[11] Geiselhart F, Otto M, Rukzio E. On the use of multi-depth-camera based motion tracking systems in production planning environments. Procedia CIRP 2016;41:759-764.

[12] Wang Y, Zhang S, Yang S, He W, Bai X. Mechanical assembly assistance using marker-less augmented reality system. Assembly Automation 2018;38(1):77-87.

[13] Sandgren J. The augmented eye of the beholder. BrandTech News 2011.

[14] Gorecky D, Campos R, Chakravarthy H, Dabelow R, Schlick J, Zuhlke D. Mastering mass customization – a concept for advanced, human-centered assembly. Academic Journal of Manufacturing Engineering 2013;11(2).

[15] Korn O, Funk M, Schmidt A. Assistive systems for the workplace: towards context-aware assistance. In Gamification: concepts, methodologies, tools, and applications 2015;1936-1949.

[16] Sand O, Buttner S, Paelke V, Rocker C. smart.assembly – projection-based augmented reality for supporting assembly workers. In International Conference on Virtual, Augmented and Mixed Reality 2016;643-652, Springer, Cham.

[17] Radkowski R. Object tracking with a range camera for augmented reality assembly assistance. Journal of Computing and Information Science in Engineering 2016;16(1):011004.

[18] Wang Y, Zhang S, Wan B, He W, Bai X. Point cloud and visual feature-based tracking method for an augmented reality-aided mechanical assembly system. The International Journal of Advanced Manufacturing Technology 2018;99(9-12):2341-2352.

[19] Bader S, Aehnelt M. Tracking Assembly Processes and Providing Assistance in Smart Factories. Proceedings of the 6th International Conference on Agents and Artificial Intelligence 2014; 1:161-168.

[20] Otto M, Prieur M, Agethen P, Rukzio E. Dual reality for production verification workshops: a comprehensive set of virtual methods. Procedia CIRP 2016; 44:38-43.

[21] Dalle Mura M, Dini G, Failli F. An integrated environment based on augmented reality and sensing device for manual assembly workstations. Procedia CIRP 2016; 41:340-345.

[22] Westerfield G, Mitrovic A, Billinghurst M. Intelligent augmented reality training for motherboard assembly. International Journal of Artificial Intelligence in Education 2015; 25(1):157-172.

[23] Bonnechere B, Jansen B, Salvia P, Bouzahouene H, Sholukha V, Cornelis J, Rooze M, Van Sint Jan S. Determination of the precision and accuracy of morphological measurements using the Kinect™ sensor: comparison with standard stereophotogrammetry. Ergonomics 2014;57(4):622-631.

[24] Xu X, McGorry RW. The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. Applied Ergonomics 2015;49:47-54.

[25] Bortolini M, Faccio M, Gamberi M, Pilati F. Including material exposure and part attributes in the manual assembly line balancing problem. IFAC-PapersOnLine 2016; 49(12):926-931.