



10th Conference on Learning Factories, CLF2020

# Learning manual assembly through real-time motion capture for operator training with augmented reality

Francesco Pilati<sup>a</sup>\*, Maurizio Faccio<sup>b</sup>, Mauro Gamberic<sup>c</sup>, Alberto Regattieric

<sup>a</sup>University of Trento, Department of Industrial Engineering, Via Sommarive 9, 38123 Trento, Italy

<sup>b</sup>University of Padua, Department of Management and Engineering, Stradella San Nicola 3, 36100 Vicenza, Italy

<sup>c</sup>University of Bologna, Department of Industrial Engineering, Viale del Risorgimento 2, 40136 Bologna, Italy

\* Corresponding Author: [francesco.pilati@unitn.it](mailto:francesco.pilati@unitn.it)

---

## Abstract

The current fourth industrial revolution significantly impacts on production processes. The personalized production paradigm which distinguishes Industry 4.0 enables customers to order unique products, defined by the specific features selected. The operators involved in the manual assembly of such workpieces have to process an enormous component variety adapting their tasks from product to product with limited learning opportunities. On the other hand, digital technologies significantly evolved in the last decade and their adoption in industrial shop floors is increasingly wider. In particular, camera-based marker-less motion capture achieved a large popularity since it represents a cheap, reliable and non-invasive solution to track, trace and digitalize human movements in different environments. Considering the presented framework, this research proposes an original hardware/software architecture to assist in real-time operators involved in manual assembly processes during the training phase to support their learning process, both in terms of rate and quality. A marker-less depth camera captures human motions in relation with the workstation environment whereas an augmented reality application based on visual feedback guides the operator through consecutive assembly tasks during the training phase. An experimental campaign is performed at the Learning factory of the Digital production university laboratory to validate the proposed architecture compared to traditional paper-based instructions provided for trainings. A real industrial case study is adopted to test and quantitatively evaluate the benefits of the developed technology compared to traditional approach in terms learning rate.

© 2020 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer-review statement: Peer-review under responsibility of the scientific committee of the 10th Conference on Learning Factories 2020

**Keywords:** operator training, aided assembly, augmented reality, depth camera motion capture, learning factory, real-time feedback, error detection, industry 4.0, digital manufacturing

## 1. Introduction and literature review

Today Industry 4.0 era is distinguished by shortened product life-cycles with a huge number of variants defined by the specific customer needs [1,2]. This results in very limited learning opportunity for human operator involved in assembly processes of such products [3]. These workers are typically involved in training periods of even shorter duration to become familiar with the novel product variants to assembly [4,5]. Augmented Reality (AR) techniques can be of help to reach this goal. AR deals with the enrichment of the real world by virtual overlays and its main ability is to show virtual object and/or instructions directly in the operator field of view [6,7]. However, the most relevant disadvantage is represented by the fact that no feedback is provided to the operator concerning the correctness of the performed action. This disadvantage represents a severe drawback of such solutions which make them not adequate to get a complete feedback loop for aided assembly both during learning and processing phases of industrial manufacturing processes [8,9]. To overcome this limitation recent literature contributions are moving toward the development of motion capture (MOCAP) devices to accurately evaluate the operator interaction with the surrounding environment. Aim of MOCAP is the accurate digitalization of the movements performed by humans monitoring the evolution over time of the position and orientation of the different limbs in a common reference system with the opportunity to provide him a real time feedback [10-12]. During the last few years, several researchers started to conduct pilot studies to implement AR techniques in production lines to improve manufacturing and assembly processes [13-15].

Starting from this scenario, this paper proposes an original hardware/software architecture (HW/SW) to assist and guide in real-time human operators involved in manual assembly processes. A markerless depth camera conceived for MOCAP purpose tracks human motions in the workstation environment. This MOCAP technology is configured to focus on the upper body of the operator and on his hands in particular. To get a remarkable accuracy of upper limbs tracking the monitored area is limited to a workbench equipped with bins for component storage. An AR solution assists in real-time the human operator during the performance of consecutive assembly tasks. In particular virtual control volumes are displaced on the workbench, around the component bins in particular, and they are superimposed on a monitor along with the RGB video to be shown to the operator. A visual feedback indicates whether the worker performed an incorrect picking action interacting with a misleading bin.

Concerning the learning process of manual tasks, the first definition of the learning curve is from the '30s for aircraft manufacturer observation that the hours needed to assemble airplanes decrease the more planes are assembled [16]. Indeed, the rate at which learning take place is not random and it is possible to accurately forecast that. Thus, a learning curve can be evaluated considering the percentage decrement of the average assembly time per unit each time the cumulative output doubles. Usually, learning curve starts with long assembly time that improves with experience, reducing the task duration after each assembly repetition or cycle. According to [17],  $t_1$  is the execution time of the assembly process at the first cycle, whereas  $t_n$  is the execution time at  $n^{th}$  cycle. The learning curve can be expressed as Eq. 1:  $t_n = t_1 \cdot n^{-b}$  (1)

The execution or cycle time of Eq. 1 decreases by a constant percentage  $\phi$  every time the quantity produced doubles.  $\phi$  is traditionally defined as learning slope or rate, with  $0 \leq \phi \leq 1$  [18]. Thus, another prevalent form of Eq. 1 is this Eq. 2:  $t_n = t_1 \cdot \phi^{\log_2(n)}$  (2)

Both  $b$  and  $\phi$  are related to the steepness of the learning curve and the slope relationship between  $b$  and  $\phi$  is represented by this Eq. 3 [19]:  $b = -\log_2(\phi)$  (3)

Furthermore, the standard time of any task is named  $t_\infty$  and it is defined as the task duration once the learning process is finished, e.g. after infinite task execution. According to [20]  $t_1$  can be estimated with a good accuracy as function of  $t_\infty$  and  $\phi$  by Eq. 4:  $t_1 = t_\infty \cdot (53.7 - 57.1 \cdot \phi)$  (4)

Thus, for  $\phi = 90\%$  the first cycle time ( $t_1$ ) is about 2.5 times longer than its standard time ( $t_\infty$ ). Accordingly, manufacturing companies shall prefer high learning rate to limit the long assembly duration during the initial production cycles of inexperienced workers during the training phase. The greater the learning rate  $\phi$ , the lower the number of assembly cycles needed to achieve the standard time is.

The aim of this paper is twofold. The first goal is to demonstrate the applicability of the proposed AR solution, able to assist in real-time the human operator during the performance of consecutive assembly tasks to reduce the learning time of inexperienced workers. The second aim is to analyse how, the learning rate  $\phi$  is influenced by the adopted training method represented by the technology used to give information to operators. Traditional paper information system is compared to the original AR solution developed. An experimental campaign based on a real

industrial application is developed to test and compare both these approaches.

The remainder of this paper is organized as it follows. The next Section 2 describes the proposed HW/SW architecture for human operator real-time tracking which results in an original AR app developed for assistance and visual feedback during the learning process of manual assembly tasks presented in Section 3. An experimental campaign is carried out to investigate the performances of the developed technology in comparison with traditional approaches to learn novel assembly processes and it is summarized in Section 4. The benefit to adopt such HW/SW architecture in comparison to traditional paper based information system is quantitatively assessed and discussed in detail in Section 5 thanks to the learning curves obtained by the experimental campaign outcomes. Finally, 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 (HW/SW) architecture able to assist in real-time the operators performing assembly activities with multiple possible variations of the executed tasks. Therefore, this architecture is conceived to be adopted to aid workers during their training phase required to learn a novel mounting sequence of a certain product or a specific variant. Concerning the hardware architecture, the markerless depth camera technology is adopted as a flexible and non-invasive solution to monitor and track the operator movements in a 3D environment. The time-of-flight technology is selected leveraging 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 markerless technology does not require the worker to wear any cumbersome suit. Finally, the wide and deep camera field of view makes this motion capture solution adaptable to several workplace layouts overcoming possible tracking occlusions (Fig. 1).

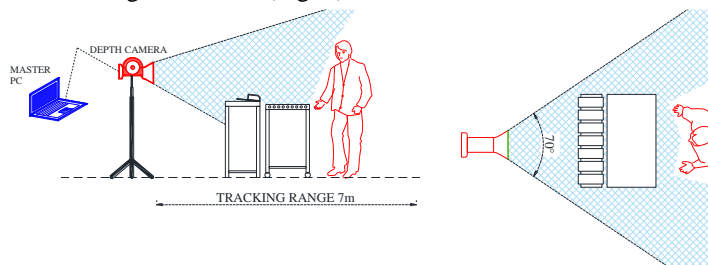


Fig. 1. Camera disposition for station layout, front (left) and top views (right side).

The proper artificial neural network encoded in the camera SDK evaluates at 30 Hz the absolute geometric 3D coordinates ( $x,y,z$ ) of each of the 20 human body joints detected by the camera. 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 or around bins ensuring a remarkable 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 representing a powerful AR solution superimposing on the RGB live video the mannequin of the operator digitalization and its interaction with the virtual CVs displaced around physical objects validating right activities and highlighting wrong actions (see Section 5).

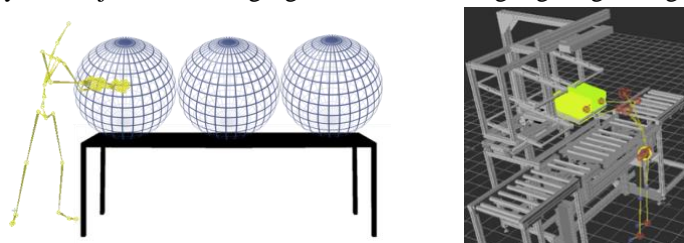


Fig. 2. Interaction between the operator hands and the CVs displaced on the workbench (left) and workstation shelves (right).

### 3. Aided assembly app

The proposed HW/SW architecture is leveraged to develop an original digital application to aid the worker during manual assembly both for training and for process assistance purposes. The proposed app relies on three consecutive phases as it follows:

1. **CVs positioning.** 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. This feature represents one of the greatest strengths of the developed app. The remarkable adaptability, scalability and absence of any physical infrastructure enable to displace the CVs wherever is needed.
2. **Expert system.** A skilled human operator executes the sequence of tasks required by the product variety to be assembled performing consecutive and ordered accesses with his hands to a subset of CVs. The app acts as an expert system which benefits from this execution storing the ordered access sequence to CVs and associating this to the considered product variety.
3. **Real-time assistance.** 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. In particular, a AR solution is conceived to guide the worker through the sequential assembly tasks highlighting on a screen the next virtual CV to visit overlaid to the physical bin containing the right component to pick. The correct picking is validated through a positive visual feedback proposing through the monitor the next CV to access. Incorrect picking actions are detected immediately and highlighted by a distinctive visual message preventing errors in the component mounting (Fig. 3).

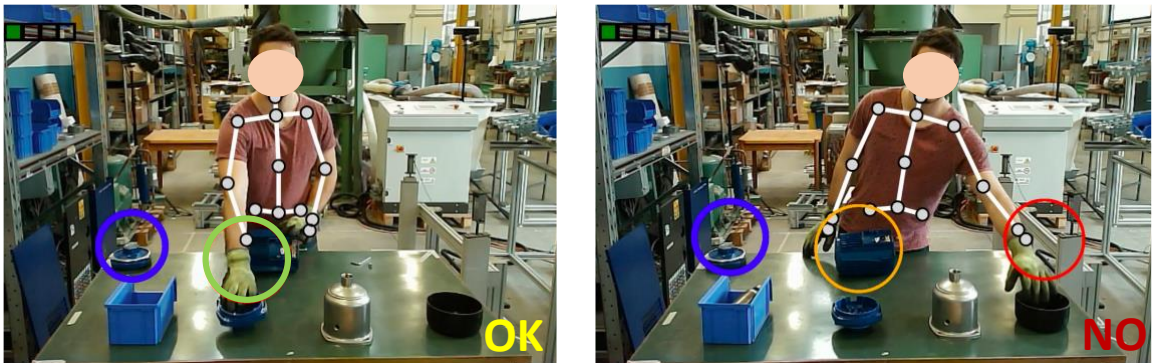


Fig. 3. Positive (left) and negative (right) visual feedback to the human operator.

### 4. Experimental campaign

An experimental campaign is performed at the Digital Production University Laboratory to test and compare the developed augmented reality architecture (ARA) for operator assistance with traditional paper-based instructions (PBI) used by operators during the training period to learn a novel manual assembly sequence. The other elements which distinguish the performed tests are listed in the followings:

- Assembly station with U-shaped layout, final product (telephone) made of 8 different components, each stored in a bin, namely 2 on the left, 3 in front and 3 on the right of the operator (Fig. 4).
- Assembly cycle performed by 12 different inexperienced operators, both female and male with various anthropometric parameters ( $o=1, \dots, 12$ ):
  - The former  $N_{PBI}=6$  of these ( $o=1, \dots, 6$ ) use paper-based instructions to get the required information and mounting sequence to perform the ordered assembly tasks of the cycle.
  - The latter  $N_{ARA}=6$  of these ( $o=7, \dots, 12$ ) use the developed AR HW/SW architecture for same purpose.
- Each operator  $o$  performs the assembly cycle 15 times ( $n=1, \dots, 15$ ). For each of these the total assembly duration is timed ( $t_{n,o}$ ). The number of assembly cycle repetition ( $n$ ) is defined according Eqs. (1-4) to reach  $t_{\infty}$  at assembly iteration end.
- For each assembly cycle  $n$  the assembly duration is calculated as the average between the different operators involved, both for PBI and ARA (Eqs. 5 and 6 respectively).

$$t_n^{PBI} = \sum_{o=1}^6 (t_{n,o} / N_{PBI}) \quad (5); \quad t_n^{ARA} = \sum_{o=7}^{12} (t_{n,o} / N_{ARA}) \quad (6)$$

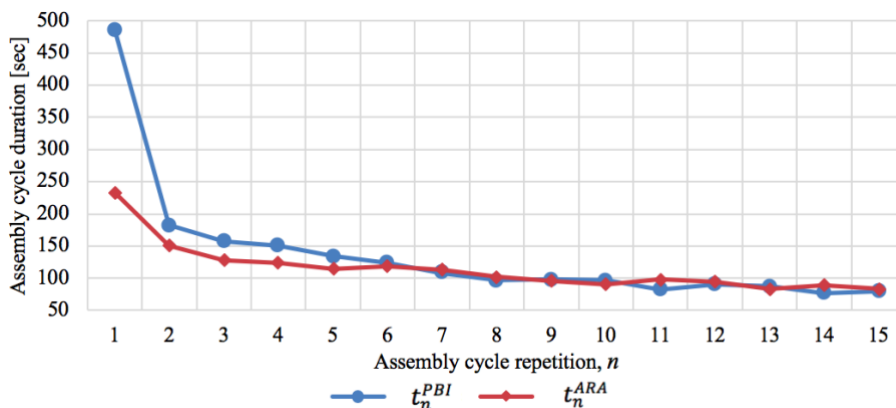


Fig. 4. Assembly station layout (left) and final product to assembly (right).

### 5. Results and discussions

This Section illustrates and discusses the experimental campaign results obtained adopting both PBI and ARA for operator assistance during the training phase. As shown in Fig. 5 (right), the duration of the first assembly cycle ( $n=1$ ) is very different comparing the two aiding solutions, e.g. PBI and ARA. Indeed, the adoption of the developed AR HW/SW architecture enables to decrease the first assembly cycle duration time of about 50%, compared to PBI (486s for PBI, 233s for ARA). Moreover, the trends of assembly cycle duration over the iterations (Fig. 5, left) suggests different learning rates  $\phi$  for PBI and for ARA aided solution, e.g.  $\phi^{ARA}=0.772$  and  $\phi^{PBI}=0.634$ . This relevant outcome suggests that to achieve the same  $t_{\infty}$  (around 85 sec.) after about 15 assembly cycles, the learning process provided by PBI is much slower than the one fostered by ARA.

A further relevant outcome to highlight deals with the 2 types of learning [20], namely “cognitive” and “motor” learning. Both types follow the traditional learning curves presented in Eqs. 1-3, but with different values for  $b$  and  $\phi$ . Cognitive learning includes decision making, following instructions, searching for cues, etc. This learning process tends to be very fast, i.e. high value for  $b$  and low value for  $\phi$ . On the other hand, motor learning is more mechanical since it includes all the physical movements needed to accomplish the task. It is a far slower learning, with low value of  $b$  and high value for  $\phi$ . At the first stages of a new learning process, the cognitive component is usually prevalent. As the assembly cycle iterations ( $n$ ) increase, the cognitive processes begin to have a lower impact and the motor processes start to dominate the learning curve, until eventually only motor learning operates. According to the results presented in Fig. 5, it is possible to state that, the usage of the developed AR HW/SW architecture during the operator training phase, let the learning process to switch from cognitive to motor first, compared to traditional PBI. This relevant outcome results in a skip of the first stage of the learning process, with a very positive impact on the duration of the first assembly cycles, e.g. for  $n=2, \dots, 5$   $t_n^{ARA}$  is lower from 15% to 20% than the corresponding  $t_n^{PBI}$ .



$n$	$t_n^{PBI}$	$t_n^{ARA}$
1	486	233
2	182	150
3	157	128
4	151	124
5	134	115
6	124	118
7	108	113
8	97	103
9	99	96
10	96	91
11	83	98
12	90	94
13	87	83
14	77	89
15	80	84

Fig. 5. Experimental campaign results: pattern (left) and values (right) of  $t_n^{PBI}$  and  $t_n^{ARA}$  for  $n=1, \dots, 15$  along with  $\phi$  and  $b$  value for both PBI and ARA as operator assistance.

$\phi$	0.634	0.772
$b$	0.678	0.381

## 6. Conclusions

This research proposes an original hardware/software architecture to assist in real-time operators involved in manual assembly processes during the training phase to support their learning process. A marker-less depth camera captures human motions in relation with the workstation environment whereas an augmented reality application based on visual feedback guides the operator through consecutive assembly tasks during the training phase. An experimental campaign is performed with an industrial case study to validate the proposed architecture compared to traditional paper-based instructions provided for trainings in terms learning rate. The obtained outcomes suggest that the duration of the first assembly cycle is much lower with ARA than with PBI assistance (-51%). Moreover, the trends of assembly cycle duration over the iterations suggests a learning rate remarkable improvement (+22%) using ARA instead of PBI aided solution.

Future research should focus on the improvement and further development of the AR HW/SW architecture considering the promising results of the presented experimental campaign. An audio feedback should be provided as alert to the operator just in case of error, to enable the operator to focus more on the assembly tasks. Of major interest is the adoption of the proposed technology for mixed-model assembly processes. ARA learning rate could probably be even more convenient compared to PBI for this production configuration.

## 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*. 103 (2019) 861–877.
- [3] 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.
- [4] Springinklee, M., & Wallenburg, C. M. Improving distribution service performance through effective production and logistics integration. *Journal of Business Logistics*, 33(2012), 309-323.
- [5] 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.
- [6] Vignais N, Miezal M, Bleser G, Mura K, Gorecky D, Marin F. Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics* (2013) 44(4):566-574.
- [7] Cohen, Y., Naseraldin, H., Chaudhuri, A., Pilati, F. Assembly systems in Industry 4.0 era: a road map to understand Assembly 4.0. *The International Journal of Advanced Manufacturing Technology*, 105 (2019), 4037-4054.
- [8] Bell, J. E., Bradley, R. V., Fugate, B. S., & Hazen, B. T. Logistics information system evaluation: Assessing external technology integration and supporting organizational learning. *Journal of Business Logistics*, 35 (2014), 338-358.
- [9] 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.
- [10] Oyekan, J., Prabhu, V., Tiwari, A., Baskaran, V., Burgess, M., & McNally, R. Remote real-time collaboration through synchronous exchange of digitised human–workpiece interactions. *Future Generation Computer Systems*, 67 (2017) 83-93.
- [11] 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.
- [12] Westerfield G, Mitrovic A, Billingham M. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education* (2015) 25(1):157-172.
- [13] Sanders, N. R., Boone, T., Ganeshan, R., & Wood, J. D. Sustainable Supply Chains in the Age of AI and Digitization: Research Challenges and Opportunities. *Journal of Business Logistics*. 40 (2019), 229-240.
- [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] 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.
- [17] Wright, T. P. Factors affecting the cost of airplanes. *Journal of Aeronautical Sciences*, 3 (1936) 122–128.
- [18] Alamri, Adel.A., Balkhi, Z.T. The effects of learning and forgetting on the optimal production lot size for deteriorating items with time varying demand and deterioration rates, *International Journal of Production Economics*, 107 (2007), 125–138.
- [19] Cohen, Y., Vitner, G., Sarin, S. Work allocation to stations with varying learning slopes and without buffers, *European Journal of Operational Research*, 184 (2008), 797–801.
- [20] Dar-El, E.M. *Human learning: From learning curves to learning organizations*. (2000) Boston: Kluwer Academic.