Mobile and adaptive User interface for human robot collaboration in assembly tasks

Dmitrii Monakhov[1], Jyrki Latokartano[1], Minna Lanz[1] Roel Pieters[1] and Joni-Kristian Kämäräinen[2]

***Abstract*— The manufacturing sector is constantly looking for more efficient ways of production. The Industry 4.0 related technologies such as augmented and mixed reality, connectivity and digitalisation as well as the current trend of robotisation have resulted a number of technical solutions to support the production in factories. The combination of human-robot collaboration and augmented reality shows good promises. The challenges in this case come from the need to reconfigure the physical production layout and how to deliver the digital instructions to the operator. This paper introduces a model for collaborative assembly tasks that uses a mobile user interface based on the depth sensors and a projector. The novelty of this research comes from the adaptivity of the user interface, as it can be freely moved between the tasks around the workstation based on the operator needs and requirements of the tasks. The ability to move projection surface is achieved by detecting the surface position using Aruco markers and computing required transformation of the projector image.**

I. INTRODUCTION

There is continuing trend towards smaller lot sizes in the manufacturing sector in Europe. As a result of increasing demand for flexibility the frequency of product changes increases as well as the number and variation of the assembly tasks. This requires higher flexibility from the employees and requires more problem solving skills and capability to assess the given situation at the factory floor [1]. A few-of-a-kind or even one-of-a kind production can only be realized if set-up times of new production processes are not very long. It is also expected that system autonomy will increase in the factory floor. This means that the systems are intelligent machines that execute high-level tasks without detailed programming and without human control [2].

As modern manufacturing systems are characterized by high complexity and an increased number of connections, which may extend outside the local boundaries, the correct and easily perceivable system architecture design becomes crucial in order to efficiently harness their full potential. For humans the situational awareness becomes a challenging task as there are multiple system components constantly changing their status. Yet, the scientific and technical development has shown that keeping the human in the loop is crucial to the system to work efficiently in non-predictive situations.

In this work we would like to present a prototype for projector-based adaptive portable user interface. The deployment of such interface can improve the overall ergonomics of the human-robot collaboration cell and it can provide additional benefits that cannot be obtained by more traditional approaches, such as improving safety awareness inside the robot cell, and providing additional help annotations to the assembly process.

II. THEORETICAL BACKGROUND

Industry 4.0 advocates the digitization of work instructions, which enables to show product specific instructions on different types of multi-media devices [3]. The most efficient way to support the operators in their work is likely via enhanced visual guidance systems [4]. The recent advances in science and technology have resulted technical achievements which have demonstrated the transformative potential of Digital and Augmented Reality (AR) guided assembly instructions. Earlier the computational and hardware limitations often dictated that these systems were deployed on tablets or other cumbersome devices [5]. In the field of Human-Machine Interaction, the User Interface (UI) plays the most important role as a mediator between the human and certain assistance systems [6]. According to the Behrendt and Strohmeier [7] it is important to improve the situational awareness of the operator. This can be achieved by combining digital assistance services, knowlegde resources, and technical devices such as UIs, microphones, tablets and speakers. Keller et al [1] also emphasised the change in the factory floor and the adaptability of the digital assistance system to the different changes. Palmarini et al. [14] notified that using AR to display information, such as robot state, progress and even intent, will enhance understanding, grounding, and thus collaboration.

The system prototype developed by [8] for AR assisted robot programming system consisted of a combination of external devices such as tablet, HDM and screen. Mueller et al [9] proposed a set-up for worker assistance system based on a combination of touch screens, cameras and laser projector. Fast-Berglund proposed [10] a concept for a Connected operator 4.0 that brings human into loop with the systems with e.g. AR headset relayed messages to the user. For example, in robotics Augmented Reality (AR) or projected instructions has been used as a human friendly interface for human-robot collaboration (HRC) [11], [12], [13]. The previous work of Hietanen et al [13] demonstrated a concept for a 3D projector-based user interface for human-robot collaboration task in mid-heavy assembly.

Palmarini et al [14] developed an AR-HRC system designed to provide context-awareness for improving human safety has been developed and tested on a pick-and-place

---

*This work was not supported by any organization

[1] Tampere University, Faculty of engineering and natural sciences, Tampere, Finland `albert.author@papercept.net`

[2] Tampere University, Faculty of Information Technologies and Communication, Tampere, Finland `b.d.researcher@ieee.org`

task. Matsas et al [15] focused on two types of techniques of safe collaboration that do not interrupt the flow of collaboration as far as possible, namely proactive and adaptive. The proactive collaboration case used audio and visual cognitive aids, which the user received as dynamic stimuli in real time during collaboration and to information enrichment for the collaboration. In this case the adaptive techniques referred to the robot; according to the first one of them the robot decelerates when a forthcoming contact with the user is traced, whilst according to the second one the robot retracts and moves to the final destination via a modified, safe trajectory, so as to avoid the human.

According to Evans et al [5] latest technical development in e.g. AR system such as HoloLens show promises, yet there is still technical challenges such as tracking accuracy, before the device is ready for deployment in a factory assembly setting. However, one of the challenges in the design of the UIs are usually designed to handle a unimodal input command (via touch screen, keyboard or mouse) and to present a feedback in a visual way. The second challenge is that they require specific devices the user must carry with him or her. The third challenge is that the UIs very rarely adapt to the physical dimensions of the user.

III. RESEARCH OBJECTIVES AND RESEARCH SETTING

The main objective of this research is to provide a contribution to the research needs identified in the literature review. In particular the objective of this research is to provide a solution for a mobile User Interface for the re-configurable production system that fits to different sizes of users without burdening the user with hand-held devices or touch screens. In this work we have scaled up the safety and UI system to the industrial scale human-robot collaboration work station and extended the solution for supporting the movement of the UI around the workstation.

The case study used in this demonstration is mid-heavy assembly of a diesel engine. The assembly consists of tasks distributed to the human or to the robot. The tasks considered in this case are pick an place tasks, where the kits are delivered to the assembly cell. The forms of collaboration are co-existence, synchronisation and cooperation [16].

The work reported in this paper focuses on the realisation of the User Interface and technical aspects to allow replication and extension of this work by others. The previous work by Hietanen et al [13] considered the operator safety aspect with greater details. In general, the novelty of this work comes from the fact that this UI solution can be adapted to the changes in the layout up to degree. However, at this stage of development, it is not yet completely in-line with the European Machine Directive [17] and robotics safety standards.

IV. IMPLEMENTATION

In this section we describe the procedure for calibrating our projector environment in collaborative work spaces. Our work environment consists of depth/RGB sensor, standard DLP projector and a predefined surface, onto which we would like project the interface. The sensor and the projector are installed on a stand above the workspace and are aimed vertically towards the floor. The process of UI generation can be split into the following steps:

- Defining Projector-Camera Homography
- Detecting the UI surface inside the projector space
- Finding orientation of the surface
- Displaying UI on the surface
- Finding Surface-Camera Homography for monitoring interactions with the interface

First, we need to find transformation from the RGB image of the sensor to the projector space [18], [19]. Projector space defines a rectangular area of the workspace, in which projector can generate images. By detecting corners of the projector space, we can define homography transformation matrix $\mathbf{H_p}$ from the original image to the projector space.

It is important to note that this homography changes with the height of the UI surface, and has to be recomputed if the user wants to use surface of different height. To detect corners of the projector space, we generate image with Aruco markers [20] at the corners, display the image using DLP projector, put our surface under each of the markers and detect the markers in the sensor's image. The code for detecting Aruco markers is implemented in OpenCV library, which was used for this project. This procedure needs to be done once and computed homography can be stored and reused in consequent operations as long as the camera-projector setup stays unaltered.

In the next step we would like to detect the location of the surface used for displaying the UI. Originally, we have tried to use contour detection methods for finding the surface. To do that we take the depth image of the projector space and detect all available contours in the area. Then we iterate over the contours to find the most suitable one in terms of area, perimeter, and shape. Unfortunately, the tests have revealed that the method is not robust to occlusions of the table area, which can happen quite often since it is natural to assume that the user would do his tasks near the UI's surface.

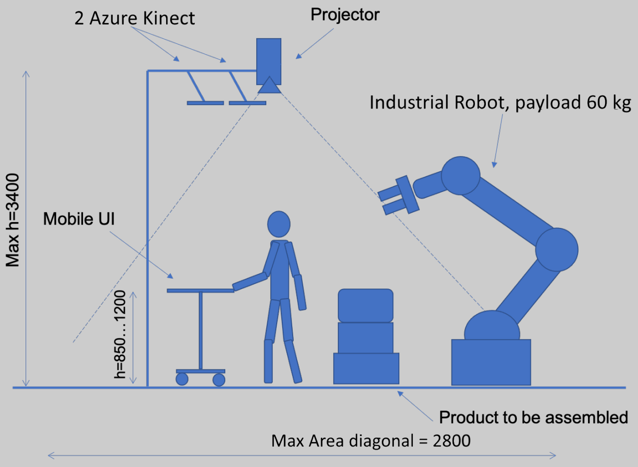

Fig. 1. Simplified picture of the system setup

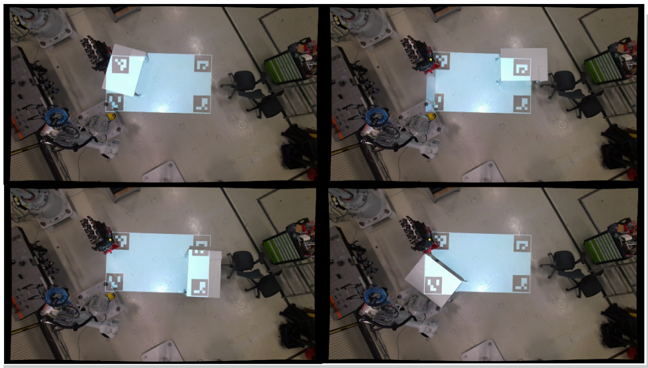

Fig. 2. Detection of corner points for camera-projector plane homography

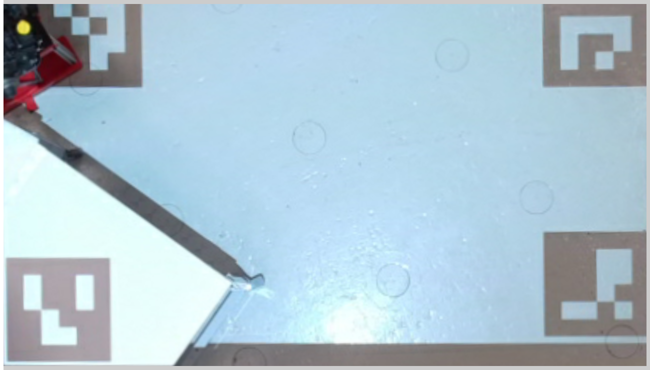

Fig. 3. Example of projector plane

Alternative method uses previously mentioned Aruco marker to define position and orientation of the surface. The marker is printed and placed at the corner of the surface with it's sides aligned with the sides of the surface. In case of rectangular surface, we can compute the ratio between the surface's side length and the marker's side length. By scaling marker's corner points with the appropriate ratios, we can find table's corners. While the method can still fail with occlusions of the marker, it is much faster and more robust than contour-based methods. It is also possible to use two markers to define rectangular areas for projections.

After the UI surface is defined, we simply find an Affine transform from the corners of the projector plane to the corners of the projection surface. To display user interface, we draw UI elements at projector's resolution and then apply the Affine transform.

Finally, we need to find transformation for monitoring interactions with the UI. To do that we find the inverse affine transform from corners of the projection surface to the corners of the screen A_T . The final transformation H_m will take the form of:

$$H_m = A_T H_p \tag{1}$$

This representation of the UI surface area is independent from the positioning of the surface, which makes it easy to compare it with the template surface area that defines the locations of the interaction elements, the same approach as in [13]. At the initialization step, we take a snapshot of the depth image in this representation, which will work as a template for registering future interactions with the UI. At each timestep, we check whether the difference between the current depth values and the template depth values is larger than some predefined threshold. If this threshold is reached, the system sends a signal that the UI element is being interacted with by user. This signal can be then handled by a different module of the system.

V. TEST SETUP

The method was tested at Tampere University HRC Pilot Line environment, where one of the demonstration tasks is a human-robot collaboration task in mid-heavy assembly. The pilot line consist of two industrial robots with a payload of 40kg and 60kg. The reach of the robots is respectively 2.55 meters and 2.05 meters making the collaborative working area size to vary between 10m2...30m2, depending on the used setup. Robots are equipped with automatic tool exchange system that provides variety of grippers for

Fig. 4. Final projection of the UI

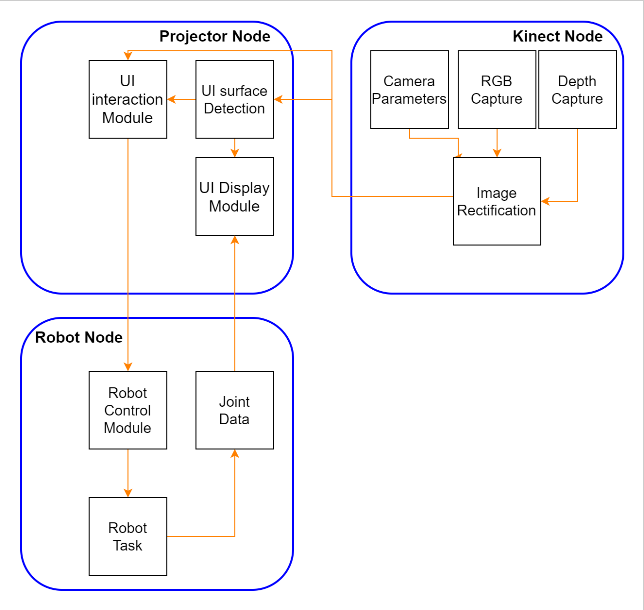

Fig. 5. Interaction between software modules

Fig. 6. Test setup

different assembly and material handling tasks. All tools can be used by both robots to increase the flexibility between different assembly needs. Robots are communicating with the projector system via ethernet connection using a ROS communication module running on robot controller. A stand is installed near the robot to hold sensors and the projector.

For the testing, A standard DLP projector and Azure kinect sensor were installed on a stand 3.5 meters above the ground in close proximity to each other. Both projector and the sensor were connected to the computer workstation, that establishes connection between the devices through ROS interface. The main three ROS nodes used in our architecture are Kinect node, which collects and transforms RGB and Depth data from the Kinect sensor, Robot node, which sends the information about the robot (status of the robot, data on the robot's joints) and can receive start/stop signals for the current operation, and Interface Node, which displays the interface and checks interactions with the UI elements. ROS is based on subscriber/publisher model in which read and write channels are processed separately from each other.

For our architecture, Kinect node provides depth and RGB data of the setup, which is continuously monitored by UI node. UI node implements methods described in previous chapter to detect UI surface and project interface onto it. When user interacts with the UI element, the Node continuously checks for a possible interaction, in case of which it would publish a message to the specified channel, for example it can be start/stop message for the robot program. The camera-projector homography of the setup is computed beforehand, and it's homography matrix is passed as an input parameter of the Interface Node. The demonstrator is set in the indoors environment under evenly distributed lighting conditions.

Fig. 7. Assessment of projection accuracy with checkerboard pattern

The experiments have shown that the setup performs well under the given conditions. The accuracy of the setup was checked with the checkerboard pattern and the average divergence error was assessed to be around 0.5-1cm, which is small enough for robust operation of the interface. The code for detecting projection surface and computing the homography was profiled, and the average runtime for the detection-projection code was equal to 45ms. This shows that projection surface detection can be run in real time, making it possible to freely move the projection surface during the experiment.

VI. DISCUSSION

While the overall experiment was a success, we would like to note some limitations of the setup from both hardware and software aspects.

Azure Kinect is continuous wave time-of-flight sensor, which requires a warm-up period before the depth measurements become stable [21] . For Azure Kinect this period should be no less than 60 minutes, which can be inconvenient for a realistic working setup. On the other hand, the standard deviation of the measurements during this process does not change considerably and depth values gradually drift towards the true depth value. Since we only check for relative change in depth values to detect interactions, there is a possibility to compensate for this drift with autoregressive moving average models. We may perform analysis for the compensation procedure in our future work.

An average DLP projector operates in range from 1 to 10 meters and with a throw ratio of 1.4-2.5. At the projector's height of 3.5 meters, the expected size of the projector plane would be around 2.5x1.4 meters. The area for the projection surface itself would be even smaller, since the surface would be higher than the ground level. While it is possible to install projector higher, two things should be taken into consideration. First is the luminosity constraint. The overall brightness of the image follows inverse square relationship with the distance. As an example, if we want to double the distance to the projector plane while keeping the same level of luminosity, the brightness level of the projector lamp should be quadrupled, making the setup quite costly. The second constraint is the resolution of the projector. By increasing the distance to the projector plane, we reduce the overall number of pixels the surface is covered with. While this may not be the problem with simple graphical elements such as buttons, the quality of text, images of videos can be greatly degraded after the increase in height. Alternative method of increasing the projection area would be stacking multiple projectors and syncing them together in one setup, which is the focus of our current research.

Lastly, let's discuss the limitations of the described UI projection procedure. In the described method, we predicate that the UI surface lies on a predefined plane parallel to the ground. This means that user cannot adjust the height of the UI surface without recomputing the projector plane homography for the new height level. UI planes that are angled towards ground can also cause larger re-projection errors, reducing the overall usability of the UI. The method also excludes multi-level planes and curved planes from possible candidates for UI surfaces. To take these variants into account, we would need more sophisticated model of the projection than a simple homography transformation, which would take into account the change in surface's depth. Since Kinect can provide us exact depth variations, deploying such model should be possible and the topic will be researched further.

VII. CONCLUSIONS

In this paper we introduced the new development for our Collaborative Assembly with Vision-Based Safety System case, which was the introduction of the movable user interface for the human robot collaboration task. The movable projector based UI allows the user to define the location of the UI based on the physical needs of the operator him/herself and the task he/she is currently doing. The technical maturity of the UI is now in such level that the user studies with the test case can be launched to test both the technical feasibility and robustness, and user friendliness. The follow-up research will naturally take two directions, where the main development path is the technical feasibility and scale-up of the system to include two robotics cells. The second path is the user side, which will focus on defining the suitable types of virtual instructions that can be projected to the UI.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825196.

REFERENCES

[1] T. Keller, C. Bayer, P. Bausch, J. Metternich, Benefit evaluation of digital assistance systems for assembly workstations, 52nd CIRP conference on Manufacuturing Systems, Procedia CIRP, 440-446, 2019

[2] R. Rosen, G. von Wichert, G. Lo, K.D. Bettenhausen,About the Importance of Autonomy and Digital Twins for the Future of Manufacturing, IFAC, Volume 48, Issue 3, 2015, Pages 567-572

[3] Romero D., Bernus P., Noran O., Stahre J., Fast-Berglund Å. (2016) The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation Towards Human-Automation Symbiosis Work Systems. In: Nääs I. et al. (eds) Advances in Production Management Systems. Initiatives for a Sustainable World. APMS 2016. IFIP Advances in Information and Communication Technology, vol 488. Springer, Cham.

[4] A. Syberfeldt, O. Danielssn, M. Holm, L. Wang, Visual Assembling Guidance Using Augmented Reality, Procedia Manufacturing 1, 2015

[5] G. Evans, J.Miller, M.Iglesias Pena, A.MacAllister, E. Winer, Evaluating the Microsoft HoloLens through an augmented reality assembly application, Proc. SPIE 10197, Degraded Environments: Sensing, Processing, and Display 2017, p.8

[6] K. Josifovska, E.Yigitbas, G. Engels, A Digital Twin-Based Multimodal UI Adaptation Framework for Assistance Systems in Industry 4.0. In: Kurosu M. (eds) Human-Computer Interaction. Design Practice in Contemporary Societies. HCII 2019. Lecture Notes in Computer Science, vol 11568. Springer, Cham. https://doi.org/10.1007/978-3-030-22636-7_30.

[7] W. Behrendt, F. Strohmeier, Orchestration and Situation Awareness in an Assistance System for Assembly Tasks, S. Ratchev (Ed.): IPAS 2020, IFIP AICT 620, pp. 332–343, 2021.

[8] N. Andersson, A. Argyrou, F. Nägele, F. Ubis, U. Esnaola Campos, M. Ortiz de Zarate, R. Wilterdink, AR-Enhanced Human-Robot-Interaction - Methodologies, Algorithms, Tools, Procedia CIRP 44 (2016) 193-198

[9] R. Mueller,M. Vette-Steinkamp, L. Hoerauf, C. Speicher, A.Bashir, Intelligent and Flexible Worker Assistance Systems Assembly Assistance Platform for Planning Assisted Assembly and Rework as Well as Execution of a Worker-Centered Assistance, Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 2: HUCAPP, pages 77-85, 2018

[10] Fast-Berglund Å., Romero D., Åkerman M., Hodig B., Pichler A. Agent- and Skill-Based Process Interoperability for Socio-Technical Production Systems-of-Systems. Advances in Production Management Systems (APMS) 2020. IFIP Advances in Information and Communication Technology, vol 592. Springer, Cham.

[11] S. Makris, P. Karagiannis, S. Koukas, A.-S.Matthaiakis, Augmented reality system for operator support in human–robot collaborative assembly, CIRP Annals Volume 65, Issue 1, 2016, Pages 61-64

[12] R-J. Halme, M. Lanz, J. Kämäräinen, R. Pieters, J. Latokartano, A. Hietanen, A., Review of vision based safety systems for human-robot collaboration, CIRP CMS 2018, Procedia Manufacturing, Stockholm, May 2018

[13] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, J-K. Kämäräinen, AR-based interaction for human-robot collaborative manufacturing, Robotics and Computer-Integrated Manufacturing vol 63, 2020, p. 18

[14] R.Palmarini, I. Fernandez del Amo, G. Bertolino, G. Dini, J.A. Erkoyuncu, R. Roy, M. Farnsworth, Designing an AR interface to improve trust in Human-Robots collaboration, Procedia CIRP 70 (2018) pp. 350-355

[15] E. Matsas, G.-C. Vosniakos, D. Batras, Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality, Robotics and Computer -Integrated Manufacturing, Vol 50, 2018, pages 168-180
I

[16] A.A.Malik, A.Bilberg, Developing a reference model for human–robot interaction, International Journal on Interactive Design and Manufacturing (IJIDeM), 2019, p.7

[17] Machinery Directive, official reference number 2006/42/EC, 2006, p. 87

[18] R. Hartley, A. Zisserman. Multiple View Geometry in Computer Vision. 2nd ed. Cambridge, UK; Cambridge University Press, 2003. p33

[19] Kämäräinen, J., Paalanen, P. . Experimental study on fast 2D homography Estimation from a few point correspondences . Lappeenranta University of Technology.(2009)
[20] F. Romero-Ramirez, R. Muñoz-Salinas, R. Medina-Carnicer(2018). Speeded Up Detection of Squared Fiducial Markers. Image and Vision Computing. 2018
[21] M. Tölgyessy; M. Dekan;L'. Chovanec;P. Hubinský Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2. Sensors 2021, 21, 413. https://doi.org/10.3390/s21020413