Combining Interactive Spatial Augmented Reality with Head-Mounted Display for End-User Collaborative Robot Programming

Daniel Bambušek, Zdeněk Materna, Michal Kapinus, Vítězslav Beran and Pavel Smrž

***Abstract*— This paper proposes an intuitive approach for collaborative robot end-user programming using a combination of interactive spatial augmented reality (ISAR) and head-mounted display (HMD). It aims to reduce user's workload and to let the user program the robot faster than in classical approaches (e.g. kinesthetic teaching). The proposed approach, where user is using a mixed-reality HMD – Microsoft HoloLens – and touch-enabled table with SAR projected interface as input devices, is compared to a baseline approach, where robot's arms and a touch-enabled table are used as input devices. Main advantages of the proposed approach are the possibility to program the collaborative workspace without the presence of the robot, its speed in comparison to the kinesthetic teaching and an ability to quickly visualize learned program instructions, in form of virtual objects, to enhance the users' orientation within those programs. The approach was evaluated on a set of 20 users using the within-subject experiment design. Evaluation consisted of two pick and place tasks, where users had to start from the scratch as well as to update the existing program. Based on the experiment results, the proposed approach is better in qualitative measures by 33.84 % and by 28.46 % in quantitative measures over the baseline approach for both tasks.**

Fig. 1: Spectator's view of the collaborative workspace, with projected user interface, extended by virtual objects seen through the HMD. Example of setting program parameters using the HMD gestures.

I. Introduction

As industrial collaborative robots are getting more affordable, it is likely that small and medium enterprises (SMEs) will soon adopt such robots in order to increase productivity. However, in such enterprises, production batches are smaller and products may be customized for a specific contract. This requires reprogramming robots for particular tasks, which could be challenging due to necessity of robot-specific knowledge. Thus it would be beneficial to enable ordinary-skilled worker to program these robots easily. Therefore, we created a prototype of a human-robot collaborative workspace – the ARCOR [1], which presents a novel approach to programming robots based on cognition, spatial augmented reality and multimodal input and output[1].

This work extends our previous solution that was dependent on a robot's presence when programming the workspace and was able to convey 2D visualization only. Integration of the head-mounted display (HMD) adds the possibility to quickly and easily visualize 3D information as e.g. pick and place positions[2] (see Figure 1). Moreover, it has a potential to at least partially eliminate the problem which occurred during the experiment from [1] where users had troubles with orientation in individual programs of the ARCOR system. Regarding the end-user programming, we focused mainly on simplifying the *pick&place* task, lowering its time to completion and the user task load.

The presented extension of the ARCOR also addresses the case of SMEs where there are more workspaces than robots. Robots are then moved between them in order to work on a workspace-specific task. To minimize enterprise losses inflicted by robot's idle time, it would be advantageous to enable workers to program the workspace even though the robot is currently working elsewhere.

II. Related Work

Recently, variety of solutions allowing end users to program robots based on AR were published. Those were based on a handheld device [2]–[4], a HMD [5], [6] or a camera-projector solution [7]–[9]. When designing the AR interface, perceptual issues as e.g. a limited field of view, a depth ordering and occlusion introduced by the selected technology and used method has to be taken into account [10]. Despite the above-mentioned problems, the AR has potential to improve HRI. For instance, it could help to avoid context switches which are normally inevitable when the user has to observe the real environment and the robot as well as the video interface [5]. Another usage could be to convey the robot's intents, especially for appearance-constrained robots [6]–[8] not able to convey those by other means.

Nowadays, spatial augmented reality (SAR) seems to be a highly promising method enabling users to interact

All authors are affiliated with the Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations, Bozetechova 1/2, Brno, 612 66, Czech Republic. Contacts: $bambusekd, imaterna, ikapinus, beranv, smrz@fit.vutbr.cz$

[1] github.com/robofit/arcor

[2] github.com/xBambusekD/ar2cor

with the robot within the task-context. For instance, its use was investigated to program a mobile welding robot [11] or in a long-term study focused on projecting assembly instructions [9]. In contrast with handheld devices, SAR has following advantages: both hands are free, projection is visible by anyone, no physical load caused by need of holding the device. On the other hand, it cannot provide free-space 3D visualization.

For unconstrained 3D interaction, HMD with integrated gesture recognition and self-localization capabilities could be used, as e.g. Microsoft HoloLens, which was a first self-contained and un-tethered device of this type. The existing solutions based on HoloLens HMD include functionality as e.g. setting of trajectory waypoints [12], previewing robot motions [13], [14], or programming of a simple pick and place task [15]. In various experiments, interfaces based on HoloLens were in many aspects (task completion times, intuitiveness, physical effort) found superior to 2D interfaces [13], [16] or to kinesthetic teaching [12]. However, for robotic applications, HoloLens limited scanning accuracy of 1-2 mm and precision of 3-5 mm [15] has to be taken into account. Moreover, interfaces has to be designed with its narrow field of view (FoV) in mind. Although the usage of HMD similarly to SAR frees users' hands, there is question of its long-term use suitability: perceived discomfort, or possible health risks.

In our approach, the HMD is used as an extension to the existing ISAR-based (interactive SAR) user interface, where it aims to provide means for effective 3D interaction (instead of kinesthetic teaching) and visualization (which was previously not possible at all). Up to our knowledge, this unique combination of the two AR techniques was not so far described in the literature. It enables us to overcome shortcomings of particular modalities and provides seamless interactive environment for letting unskilled users to program complex robotic tasks.

III. Proposed Mixed Reality Interface

The baseline approach of the end-user robot programming uses the ISAR in a combination with kinesthetic teaching (ISAR-KT). We use the kinesthetic teaching only for setting the target pose, while the robot computes trajectory to it by itself according to the current state of the workspace. In order to fulfill outlined goals (remove robot dependency, reduce programming completion time and user's task load), we replaced the kinesthetic part with the HMD. Thus, we are proposing an approach that uses a combination of the ISAR with HMD (ISAR-HMD)[3].

A. Setup

The ARCOR setup, that we created, consists of a projector, which projects a user interface onto a touch-enabled table that forms the ISAR, two Microsoft Kinect sensors, two speakers placed beneath the table and a robot. As a demonstrator of a collaborative robot, the PR2 is used. For a computational power, three processing units (Intel NUC) connected into a wired network, are used. The base setup is described in more detail in [1].

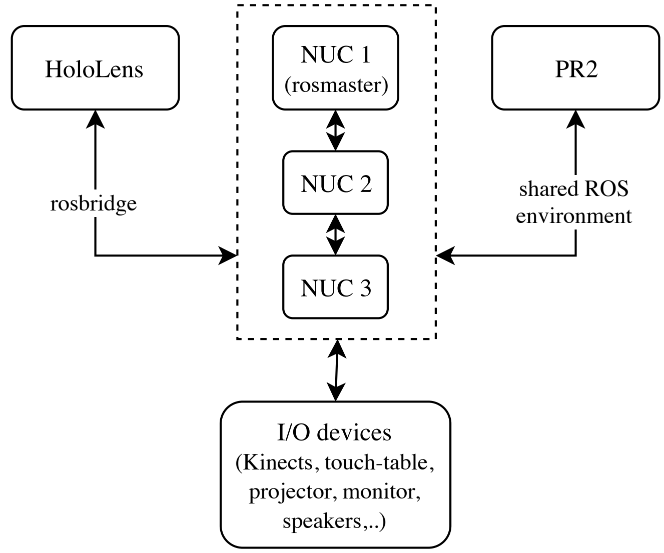


Fig. 2: ARCOR architecture. Intel NUCs are used as processing units, communicating together via local network. The whole system runs on Ubuntu along with ROS. Microsoft HoloLens communicates with the ROS environment via the rosbridge API.

In order to overcome the inherent limitations of the ISAR-based solution, a HMD (Microsoft HoloLens 1) was integrated, which serves for visualization and interaction in 3D, whereas ISAR could still be used for tasks where 2D interaction is sufficient. The HoloLens communicates with the main processing unit, which runs on the Ubuntu 14.04 along with ROS Indigo, via the rosbridge API (see Figure 2). Since there are individual sensing units – Kinects, robot, HoloLens – they all need to be calibrated with respect to a common point (corner of the table). The calibration procedure is based on detection of a known AR marker.

Projected user interface offers various widgets. Most crucial is a program list, where all stored programs are displayed. User can either edit parameters for already set program or template selected program, which will create a new instance of it with empty parameters. Each program can be composed of multiple blocks, where each block has a set of instructions for which the parameters, like pick position, object type, place position, etc. needs to be set.

B. Problems to Solve in the Pick&Place Use Case

We focused mainly on simplifying the *pick&place* learning procedure. More specifically, on a scenario where user wants to set the robot to pick an object from the feeder and place it on the table.

In order to properly program this task in the ARCOR setup, the user has to set three main parameters – *detection position*, *object type* and *place position*. Due to the HoloLens limited scanning accuracy (1-2 mm), its precision (3-5 mm) [15] and a possible inaccurate user input,

[3] Video of the proposed approach: https://youtu.be/MNXhqpFBy9Y.

robot's cognitive abilities (attached forearm cameras) are used, which are able to find the object of specified type from the detection position in order to determine precise picking position.

Since the proposed solution (ISAR-HMD) is aiming on elimination of robot's presence, the user uses only touches of the touch-enabled table and gestures of the HMD for interaction with the system. Without the kinesthetic teaching and with the goal of keeping the ISAR-HMD as simple as possible, we needed to solve:

- How to efficiently select the object type to be picked up.
- How to set the detection position, from which the robot will be able to detect and pick up the object.
- How to set the place position and its rotation.

C. Solution to the Pick&Place Use Case

Based on the results from [14], a heading-based selection, where user is using his gaze for targeting (indicated with virtual cursor) and a hand for the selection gesture (HoloLens Air tap – equivalent of mouse click), is used.

For a sake of efficiency (lowering the number of actions the user has to take), setting the picking instruction (named *pick from feeder* in the ARCOR system) and setting the placing instruction (named *place to pose* in the ARCOR system) is tied together to form a fluent procedure.

All visible objects in the scene are detected and registered (Kinect sensors), making them interactive for the HMD. While the user is gazing at such detected object during setting the detection position, visual feedback – in a form of virtual robot gripper rendered with $0.3\,m$ offset from the HMD's cursor – is provided. This gripper, which is automatically positioned against the side of the object the user is looking at, is indicating current detection position directly in the scene. When colored green, the robot will be able to detect and pick up the object, when colored red, the robot will not be able to do so.

Final stage of setting the object type and the detection position is merged into one action – HoloLens Air tap gesture (equivalent of mouse click) on desired object in feeder. As the object is detected, the system automatically recognizes and saves the type of it, and as the HMD is calibrated with respect to the ARCOR system, the position and rotation of the virtual gripper is transformed to the ARCOR coordinate system and saved as the detection position.

Since the setting of the *pick from feeder* instruction is tied up with the setting of the *place to pose* instruction, a virtual object of the type the user selected in the previous step is created and attached to the end of user's gaze in order to create the illusion of naturally picking an object from the feeder and placing it on the table (we are benefiting from the HoloLens spatial mapping abilities, where the attached virtual object can collide with the real environment). For this purpose, we chose the *click-attach-click* approach (click on the object, attach the virtual one, click on the table to release it) rather than the *drag&drop*, because the virtual object could easily lost from the user's sight or the hand tracking of the HoloLens could easily lost (because of the limited FoV for hand recognition).

After placing the virtual object onto the table inside the reach zone of the robot (visualized by the SAR projection), virtual spheres, for setting the rotation, are displayed. By dragging them, the rotation is set.

The procedure of setting the *pick&place* program using the ISAR-HMD can be summarized into following steps:

1) Click on the *Edit button* of the *pick from feeder* instruction in the projected interface.
2) Look at desired object placed in the feeder, position the virtual robot gripper to desired detection pose and click on it (Air tap).
3) Position attached virtual object on the table and click when satisfied with the position.
4) Adjust the rotation by dragging displayed spheres around the virtual object.
5) Click on the tick button in the HoloLens or on the *Save button* of the projected interface to confirm and save the place position.

Whole procedure of setting the *pick&place* program using the ISAR-HMD is shown in the Figure 6.

Programming procedure, when using the ISAR-KT, is similar in the projected interface related steps – 1 and 5. In the step 2, user has to physically move the robot's gripper to desired detection position. As the arm is in interaction mode, its forearm cameras are on, seeking for any visible objects. If any are visible, the robot recognizes the type of the closest one. When the user is satisfied with the set object type and detection position, he saves it using the *Save button* of the projected interface. Thus *pick from feeder* instruction is set. Learning of the following *place to pose* instruction needs to be called manually. Robot's reach zone as well as interactive bounding box representing the place position are displayed. User drags the bounding box outline and blue point situated in its corner in order to set the place position and its orientation (steps 3-4). The procedure is shown in the Figure 5.

It has to be mentioned, that ISAR projections are synchronized with HMD's virtual objects and vice versa. Meaning that the user can whenever decide, if he wants to set the place positions using the touches on the table or gestures in HoloLens. It is also possible to put aside the HMD at anytime and continue the programming using the ISAR-KT approach.

D. Main Benefits

In a scenario, where company has multiple workspaces but limited number of collaborative robots that are moved between those workspaces, it would be time consuming to edit current programs at individual workspaces, because the need of robot's presence if kinesthetic teaching is applied. However, using our solution, the robot is not needed. Workers can effectively set programs in advance anytime, without the need of stopping the production of a current batch.

Thanks to the combination of the HMD with the ISAR, others are partially able to see directly in the scene in

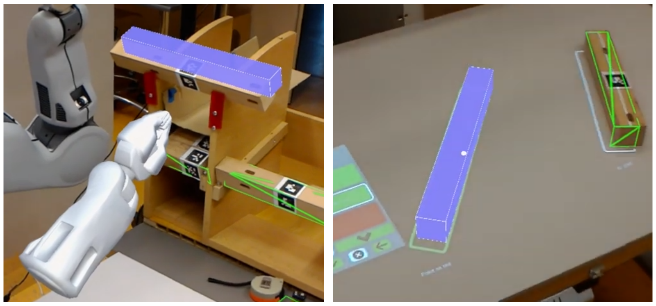


Fig. 3: Visualization of instructions with already set parameters. **Left:** Virtual gripper is rendered on a set detection position along with the virtual object of specified type (*pick from feeder* instruction). **Right:** Virtual object is rendered on a set place position (*place to pose* instruction).

realtime, what the user with the HMD put on is currently doing, which is not possible without any additional device (hand-held device, another pair of HMD, or HMD's stream). Moreover, the ISAR extends the HMD's limited FoV by 2D projections. As far as we know, no one ever combined those two augmented reality approaches.

If the program is set, the user can see a virtual gripper rendered directly on the detection position along with the virtual object, that is going to be picked, in case of *pick from feeder* instruction, or rendered virtual object on the place position in case of *place to pose* instruction (see Figure 3). This is beneficial if the user just wants to preview the program without running it. It could also positively impact the users' orientation within set programs.

We used a text-to-speech utility in order to play system related notifications, warnings and errors to users in their native language through the HoloLens embedded speakers. This could be beneficial for new users of our system.

IV. Experiment Design

In order to evaluate the proposed ISAR-HMD solution, an experiment was designed, where the solution was compared to the baseline ISAR-KT approach. Both approaches were tested on a set of 20 participants using the within-subject design methodology. Order of conditions was randomized to mitigate possible bias caused by a learning effect.

As measures, we chose a combination of three standardized questionnaires – the System Usability Scale (SUS) [17], NASA Task Load Index (TLX) [18] and the User Experience Questionnaire (UEQ) [19]. We also measured task completion times.

A. Hypotheses

As the main motivation for this work is to introduce a novel approach of teaching robots that could replace the kinesthetic teaching, we assume that our ISAR-HMD solution will be quicker, less demanding and more preferred by users than the ISAR-KT. Therefore, we set following three hypotheses:

(i) The ISAR-HMD approach is faster than the ISAR-KT approach.

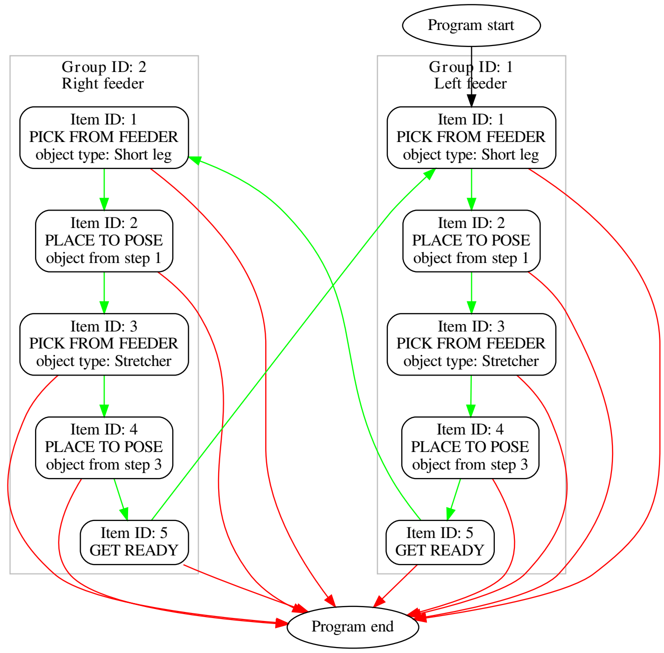


Fig. 4: Pick&Place program. The green edges represents transitions, that are triggered if current instruction was successfully executed, while the red ones represents transition of instruction's unsuccessful execution. The program is designed to run in loop, until user decides to stop it.

(ii) User task load of the ISAR-HMD approach is lower than the ISAR-KT approach.

(iii) In terms of UX, users will prefer ISAR-HMD over ISAR-KT.

B. Tasks

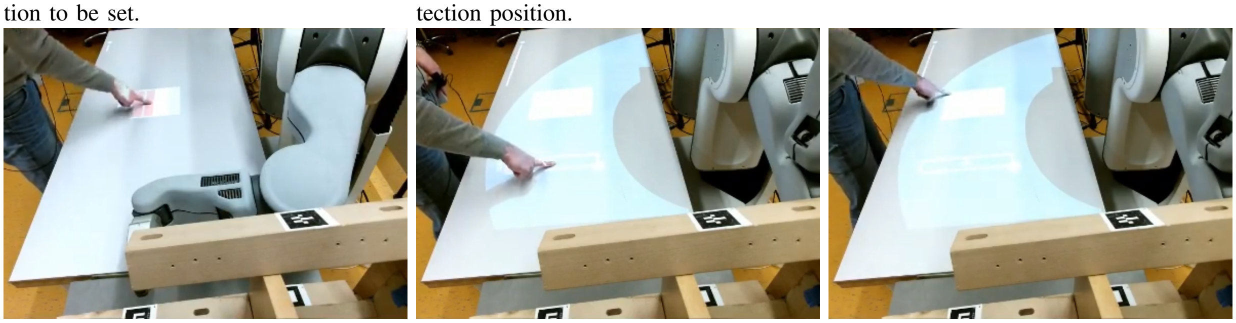
Experiment workflow of tested methods consisted of following phases: *introduction*, *training*, *first task*, *second task* and *questionnaire* ended up with *discussion*. After the user finished the workflow using one of the methods, he/she repeated it using the other one.

Within **the introduction phase**, participants got an overall idea of the collaborative robotics purpose, its related problems we are solving and a brief description of the upcoming experiment.

The training phase involved demonstrative and commented setting of one pair of the *pick from feeder* and the *place to pose* instruction, using currently tested method. In case of the ISAR-HMD method, the training phase involved getting familiar with the HMD (HoloLens). Participants went through the Microsoft's *Calibration* application – to calibrate their interpupillary distance, which can improve the quality of visuals – and the Microsoft's *Learn Gestures* application – to ensure they properly learn how to use the HoloLens gestures.

The first task consisted of setting parameters for an unset *pick&place* program. This program was composed of two blocks, the first one for picking from feeder on user's left side and the second one for picking from feeder on user's

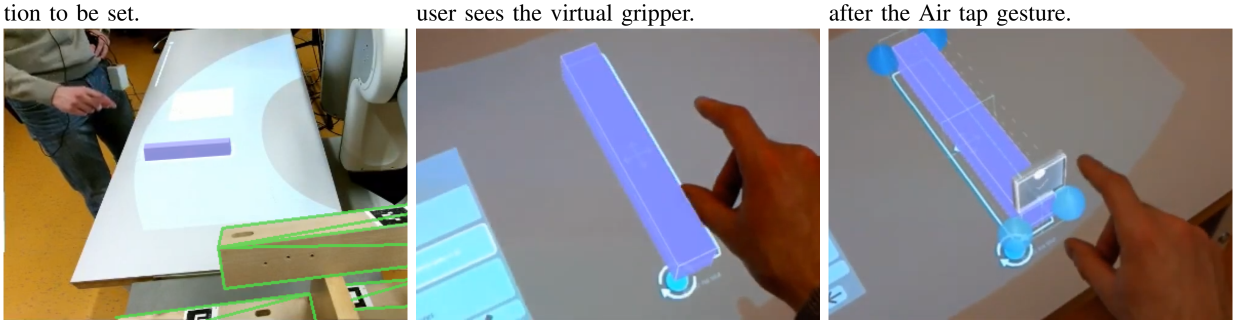
(a) User selects *pick from feeder* instruction to be set. (b) User moves the robot gripper to detection position. (c) Gripper's detection position is saved.



(d) User selects *place to pose* instruction to be set. (e) User adjusts place position by dragging it on the table. (f) The place position is saved.

Fig. 5: An example of setting the *pick&place* program using the ISAR-KT approach during the experiment.

(a) User selects *pick from feeder* instruction to be set. (b) User's first person view. While gazing, user sees the virtual gripper. (c) Virtual object snaps to user's gaze after the Air tap gesture.



(d) Place position is adjusted by user's head movements. (e) When user clicks (Air tap), virtual object snaps to the table and the rotation spheres are displayed. (f) User saves the place position by clicking on the tick button.

Fig. 6: An example of setting the *pick&place* program using the ISAR-HMD approach during the experiment.

right side. Both blocks contained four parametric instructions – two *pick from feeder* and two *place to pose*, and one non-parametric instruction – *get ready* (moves the robot's arms to their default position). The *pick from feeder* instruction takes two parameters – object type and robot's gripper position for detecting the objects in feeder followed up with picking the closest one. The *place to pose* instruction takes just object's place position as the parameter (object type is referenced from previous pick instruction). Structure of the program is shown in Figure 4. An example of completing part of this task using both tested methods is shown in Figure 5 (ISAR-KT approach) and Figure 6 (ISAR-HMD approach).

The second task consisted of editing preset *pick&place* program. Within made up backstory, we told participants that someone mistakenly set the program with wrong parameters (wrong object types, wrong place positions, overlapping place positions, etc.). Their task was to detect those instructions with wrong parameters and correct them to fulfill the assignment. The program had same structure as the program in the first task.

In order to be able to record participant's point of view and head tracking for both conditions (which was necessary for evaluation purposes), participants wore the HMD even in condition where it was not actually used by them. Moreover, this could prevent distortion of the results caused by potential discomfort from wearing the HMD, which would not be the case for forthcoming devices as e.g. HoloLens 2 (lighter, better balanced).

C. Participants

Prior the main study, the experiment design was tested out in pilot test with 2 participants. After that, 20 users participated in the main experiment. Most of the participants were IT students or faculty employees (18 male and 2 female, ages 20-31, $M = 25.00$). 13 participants never used VR/AR HMD. There were total of 12 participants reporting eye issues. 5 of them reported farsightedness, 1 reported nearsightedness and 5 reported wearing glasses or contact lenses without specifying exact eye issue. 1 reported color blindness. On a Likert scale from 1 to 5, most of participants expressed positive attitude towards new technologies ($M = 3.55, CI =< 3.19, 3.91 >$) and rather high IT skills ($M = 4.00, CI =< 3.41, 4.59 >$). On the other hand, experience with robots ($M = 2.00, CI =< 1.50, 2.50 >$) and experience with AR ($M = 2.20, CI =< 1.68, 2.72 >$) were self-assessed rather low, which could be expected to be close to the situation in the target user group (employees in SMEs).

V. Results

This section summarizes the experiment results and provides its analysis and interpretation. Regarding the task completion time measurement, intervals where participants were asking questions, technical problem occurred or when moderator had to intervene, were subtracted, in order to measure a pure task completion time. All statistical tests were done at the $5\,\%$ significance level. Data were first tested for normality (combination of D'Agostino and Pearson's tests) and based on the result, paired t-test (**T**) or Wilcoxon's signed-rank test (**W**) were used to test for the significant difference between conditions.

A. Quantitative and Qualitative Data

Both tasks were completed quicker when using the proposed solution (ISAR-HMD). Completing them both using proposed solution saved up to 153.94 seconds in average, which confirms the hypothesis (i).

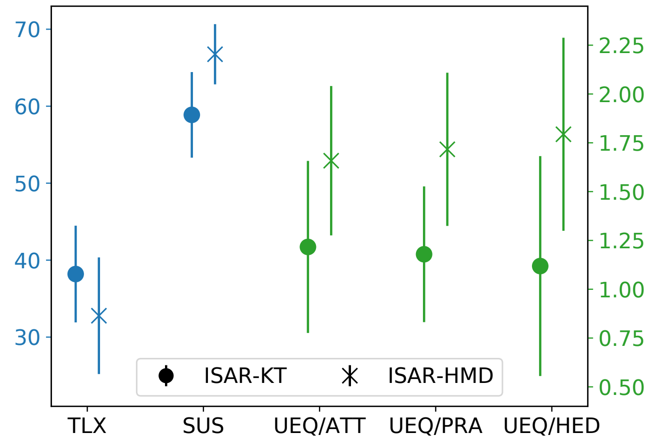


Fig. 7: Obtained qualitative measures (mean values and 95 % confidence intervals) for both evaluated conditions. The left y axis belongs to TLX and SUS, while the right one to the UEQ grouped scales.

Table I shows the results of measured metrics for both tested methods. There is a statistically significant difference for all metrics, except the TLX and majority of its subscales. The mean TLX score of the proposed ISAR-HMD solution was 32.78 which is less than 38.19 for the ISAR-KT; however, the difference is not statistically significant. Thus the hypothesis (ii) cannot be confirmed. For TLX subscales, there is a significant difference for *Temporal Demand*, corresponding to the objective measurement of task times, where ISAR-HMD was significantly faster then ISAR-KT for both tasks. Similarly to [12], the kinesthetic teaching required lower mental load and higher physical demand. However, the differences were not significant. We hypothesize, that higher mental demands for the ISAR-HMD are mainly caused by a hardware limitations of the used device, namely limited FoV for visualization (leading to increased demands on the user's spatial cognitive abilities) and for capturing gestures. Also the interface should cope with the given limitations better – e.g. it could indicate direction to the interactive elements which are currently out of the user's FoV.

The hypothesis (iii) is well supported by obtained UX-related ratings. The mean SUS rating of the ISAR-HMD approach was 66.75 which means improvement over the ISAR-KT (58.88). Figure 7 shows measured metrics – its mean values and confidence intervals – in a graph. UEQ consists of six categories, where some can be grouped together – **attractiveness** (ATT); **pragmatic quality** (PRA) that encapsulates perspicuity, efficiency and dependability; and **hedonic quality** (HED) that encapsulates stimulation and originality. According to the general benchmark [19], ratings of all three main categories of the ISAR-KT approach are ranked as *Above Average*. Ratings of the ATT and PRA of the ISAR-HMD approach are ranked as *Good*, which is one rank higher than the ISAR-KT and the HED score is ranked as *Excellent*, moving it into the top rank.

Further, we divided both quantitative and qualitative data

Measure	ISAR-KT	ISAR-HMD	T/W Value	p
SUS	$58.88; < 53.32, 64.43 >$	$66.75; < 62.84, 70.66 >$	$T(20) = 3.55$	$\mathbf{0.002}$
NASA TLX	$38.19; < 31.91, 44.48 >$	$32.78; < 25.20, 40.35 >$	$T(20) = 1.31$	0.206
NASA TLX / Mental Demand	$17.5; < 8.56, 26.44 >$	$25.83; < 13.06, 38.61 >$	$T(20) = 1.39$	0.180
NASA TLX / Physical Demand	$23.33; < 12.19, 34.48 >$	$16.67; < 7.20, 26.14 >$	$T(20) = 1.22$	0.237
NASA TLX / Temporal Demand	$41.67; < 28.64, 54.69 >$	$24.17; < 12.44, 35.89 >$	$T(20) = 2.27$	$\mathbf{0.035}$
NASA TLX / Overall Performance	$90.00; < 80.74, 99.26 >$	$86.67; < 77.68, 95.65 >$	$W(20) = 17.50$	0.546
NASA TLX / Effort	$30.83; < 19.16, 42.51 >$	$24.17; < 14.54, 33.80 >$	$T(20) = 0.94$	0.359
NASA TLX / Frustration Level	$25.83; < 16.54, 35.12 >$	$19.17; < 8.06, 30.28 >$	$T(20) = 1.51$	0.148
UEQ/ATT	$1.22; < 0.78, 1.66 >$	$1.66; < 0.78, 1.66 >$	$T(20) = 2.26$	$\mathbf{0.036}$
UEQ/PRA	$1.18; < 0.83, 1.53 >$	$1.72; < 1.32, 2.11 >$	$T(20) = 2.90$	$\mathbf{0.009}$
UEQ/HED	$1.12; < 0.56, 1.68 >$	$1.79; < 1.30, 2.29 >$	$W(20) = 16.50$	$\mathbf{0.001}$
1st task completion time (s)	$282.58; < 248.88, 316.28 >$	$196.18; < 152.84, 239.53 >$	$W(20) = 22.00$	$\mathbf{0.002}$
2nd task completion time (s)	$256.33; < 205.93, 306.73 >$	$188.78; < 145.85, 231.72 >$	$T(20) = 2.60$	$\mathbf{0.017}$

TABLE I: Qualitative measures (System Usability Scale, NASA Task Load Index and its subscales, User Experience Questionnaire which consists of three categories – Attractiveness, Pragmatic Quality and Hedonic Quality) and quantitative measures (task completion times). The data for both methods are in format "mean; and respective $95\,\%$ confidence interval". For a statistical comparison, we used paired t-test (**T**) and Wilcoxon (**W**) method.

into two parts according to following binary conditions: previous experience with HMD, presence of an eye-related health problem and order of evaluated method (whether user tested ISAR-HMD first). Differences in measures between aforementioned parts for both evaluated methods were tested using a t-test for independent samples or Kolmogorov-Smirnov's test based on normality test result. No statistical significant differences were found. Our interpretation is that HMD is suitable even for novice users without previous experience with HMD. Further, task completion times nor subjective assessment of the method are influenced by an existence of vision-related health problem or limitation, meaning that the used HMD device (HoloLens 1) does not posses problems for users wearing glasses, etc. Finally, in contrary to [16] where users rated 2D interface significantly lower after they interacted with the system using HMD, in our case no order effect was identified. This could mean that both methods (interfaces) are acceptable, likeable and roughly equally hard to learn and use.

B. General Findings

Biggest downside of our setup was probably unreliable touch-enabled table. False touches, double-clicks or undetected touches were source of frustration for most participants and probably caused the overall low ratings of the qualitative data.

Three participants struggled with positioning the robot's gripper. They were not able to rotate the arm links properly in order to find correct kinematic configuration.

On the other hand, 4 participants had troubles with learning and adopting the HoloLens Air tap gesture. Main source of such problems was caused by not having hands in HoloLens cameras detection zone. Three complained about the HMD's text-to-speech, claiming that they already know what to do and what is happening after the training phase and few set instructions from the first task. One participant reported headache after completing both tasks using the HMD.

One participant suggested that it would be better, if he could oversee all place poses at once. This suggestion needs to be further tested, because an overwhelming number of virtual objects displayed at once could cause user's confusion and inability to orientate within the program.

Interestingly, some participants preferred to use the touch table to adjust object's place position, even though they were supposed to use primarily the HMD's gestures. We also noticed few situations where participants were not able to distinguish SAR projections from HMD virtual objects. They tried to interact with those projections using the HoloLens gestures and not the touch-enabled table. This could be a good indicator that ISAR can be visually believable merged with the HMD's AR without direct distinction.

VI. Conclusions

In this paper, we presented a novel approach to the end-user robot programming using the unique combination of the interactive spatial augmented reality and the head-mounted display (ISAR-HMD). The purpose of this approach is to reduce the users' workload and to let them program the collaborative workspace faster than in kinesthetic approaches and without the need of the robot's presence.

We evaluated the proposed ISAR-HMD approach on a set of 20 participants using the within-subject experiment design, where we compared it to the baseline approach, that uses the interactive spatial augmented reality and the kinesthetic teaching only (ISAR-KT). We reached up to $33.84\,\%$ improvement in qualitative measures (SUS, NASA-TLX, UEQ) and saved up to $28.46\,\%$ of completion time of setting the *pick&place* program.

In the future work, we will focus on lowering the task load for HMD, which could be achieved by a constrained FoV-aware visualization. Another direction of the research will be further integration of ISAR and HMD. Additionally, we are going to develop a detector of UX-related events, based on combination of physiological data with data from external sensors (e.g. user pose tracking) and input data (e.g. clicks), which could be helpful for the system to automatically provide timely assistance to the user as an excessive amount of voice notifications was one of the most common complains from users in the current experiment.

ACKNOWLEDGMENT

The work was supported by Czech Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project "IT4Innovations excellence in science – LQ1602".

REFERENCES

[1] Z. Materna, M. Kapinus, V. Beran, P. Smrž, and P. Zemčík, "Interactive spatial augmented reality in collaborative robot programming: User experience evaluation," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2018, pp. 80–87.

[2] G. Michalos, P. Karagiannis, S. Makris, Ö. Tokçalar, and G. Chryssolouris, "Augmented reality (ar) applications for supporting human-robot interactive cooperation," *Procedia CIRP*, vol. 41, pp. 370–375, 2016.

[3] S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, and M. Tscheligi, "Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 179–184.

[4] S. Magnenat, M. Ben-Ari, S. Klinger, and R. W. Sumner, "Enhancing robot programming with visual feedback and augmented reality," in *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 2015, pp. 153–158.

[5] H. Hedayati, M. Walker, and D. Szafir, "Improving collocated robot teleoperation with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 78–86.

[6] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 316–324.

[7] E. Bunz, R. T. Chadalavada, H. Andreasson, R. Krug, M. Schindler, and A. Lilienthal, "Spatial augmented reality and eye tracking for evaluating human robot interaction," in *RO-MAN 2016 Workshop: Workshop on Communicating Intentions in Human-Robot Interaction, New York, USA, Aug 31, 2016*, 2016.

[8] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, pp. 241–248, 2014.

[9] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, and A. Schmidt, "Working with augmented reality?: a long-term analysis of in-situ instructions at the assembly workplace," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2017, pp. 222–229.

[10] E. Kruijff, J. E. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," in *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, pp. 3–12.

[11] R. S. Andersen, S. Bøgh, T. B. Moeslund, and O. Madsen, "Task space hri for cooperative mobile robots in fit-out operations inside ship superstructures," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 880–887.

[12] C. P. Quintero, S. Li, M. K. Pan, W. P. Chan, H. F. Machiel Van der Loos, and E. Croft, "Robot programming through augmented trajectories in augmented reality," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 1838–1844.

[13] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," *CoRR*, vol. abs/1708.03655, 2017. [Online]. Available: http://arxiv.org/abs/1708.03655

[14] D. Krupke, F. Steinicke, P. Lubos, Y. Jonetzko, M. Görner, and J. Zhang, "Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 1–9.

[15] S. Blankemeyer, R. Wiemann, L. Posniak, C. Pregizer, and A. Raatz, "Intuitive robot programming using augmented reality," *Procedia CIRP*, vol. 76, pp. 155–160, 01 2018.

[16] S. Yitzhak Gadre, E. Rosen, G. Chien, E. Phillips, S. Tellex, and G. Konidaris, "End-user robot programming using mixed reality," 10 2018.

[17] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[18] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

[19] M. Schrepp, "User experience questionnaire handbook," 09 2015.