XR-Cockpit: a comparison of VR and AR solutions on an interactive training station

Ariel Caputo
Department of Computer Science
University of Verona
Verona, Italy
cptfmr71@univr.it

Sergiu Jacota
Department of Computer Science
University of Verona
Verona, Italy
sergiu.jacota@studenti.univr.it

Serhiy Krayevskyy
Department of Computer Science
University of Verona
Verona, Italy
serhiy.krayevskyy@studenti.univr.it

Marco Pesavento
Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, Surrey, United Kingdom
m.pesavento@surrey.ac.uk

Fabio Pellacini
Department of Computer Science
Sapienza University of Rome
Rome, Italy
pellacini@di.uniroma1.it

Andrea Giachetti
Department of Computer Science
University of Verona
Verona, Italy
0000-0002-7523-6806

***Abstract*—One of the most challenging aspects of the implementation of Virtual/Mixed reality training systems is the effective simulation of real-world manipulation of the physical devices included in control interfaces like buttons, sliders, levers, knobs, etc. In this paper we describe a mockup airplane cockpit (XR-Cockpit), featuring interactive components of this kind that demonstrate the feasibility of effective simulations of device manipulation using low cost hand tracking technology and gesture recognition. Based on this system, we performed a user study to compare the effectiveness of the interaction with virtual tools using different visualization solutions: immersive VR, optical and video see-through based MR. In our study, we also checked how well it is possible to perform manipulation of real objects wearing the two video see-through solutions. The analysis of the experimental results provides useful guidelines for the design of Virtual and Mixed Reality training systems involving virtual and physical actions on manipulation devices.**

***Index Terms*—Virtual Training, Virtual Reality, Mixed Reality, Manipulation**

I. Introduction

Virtual and Mixed Reality (VR/MR) are considered enabling technologies for the Industry 4.0 paradigm and, in the next few years, are expected to play an important role in the manufacturing domain [1]–[3]. VR/MR applications are also growing in popularity in many other application domains like Healthcare or Cultural Heritage [4]–[6]. The availability of low-cost Head-Mounted Displays (HMD), finger tracking solutions and other interaction tools [7], enables the creation of novel exciting, and affordable applications.

One of these applications is virtual training, as it is possible to create virtual replicas of complex interfaces, like an industrial control panel or an airplane cockpit, allowing users to learn complex sequences of actions without the need of working on the actual tangible and often expensive devices. Furthermore, using the virtual interaction devices, it is possible to optimize the interaction design, testing interactive components' functionality and ergonomics on the virtual replica or in a mixed reality setting including only a part of the real interface without the need of realizing tangible mockups of the components with different features. However, as shown in [8], manipulating virtual devices is still a big issue in VR.

In this paper, we present an example of VR/MR simulator allowing different kinds of manipulation-based interactions and potentially useful for training purposes in an industrial context or to realize ergonomic studies. This example is a virtual airplane cockpit prototype (XR-Cockpit), actually developed in an ongoing industrial collaboration. The cockpit has several active parts that can be manipulated by the user in a way that resembles the real parts' manipulation. The implementation of these interactive parts is based on a low-cost hand tracker (Leap Motion) and a mix of simulation based on Unity framework's physics engine [9] and gesture recognition and remapping of user movements over virtual objects when the tracking accuracy does not allow for smooth physically-based interaction (e.g. precise grabbing or finger rotation).

Other than demonstrating the feasibility of a smooth interaction with its virtual devices, we also exploit XR-Cockpit to evaluate the effects of different technical solutions for the immersive visualization.

First, we investigate whether the use of MR, e.g. anchoring the virtual scene to real-world surroundings improves the interaction experience or not. Compared to VR, the use of MR HMDs avoids the risk of collisions with real-world objects and allows the realization of interface mock-ups mixing real and virtual parts, that can be useful in the interaction design. Some authors also claim that MR affords better spatial perception [10].

Second, we try to understand which HMD solution is currently the best for our task in the case of MR implementation between a camera see-through and optical see-through vision. Current MR HMDs with semitransparent displays, like Microsoft's Hololens, have a limited field of view (FOV), that affects the usability of MR application while current camera see-through solutions (we considered a setup with Zed Mini

978-1-7281-8956-7/20/$31.00 ©2020 IEEE

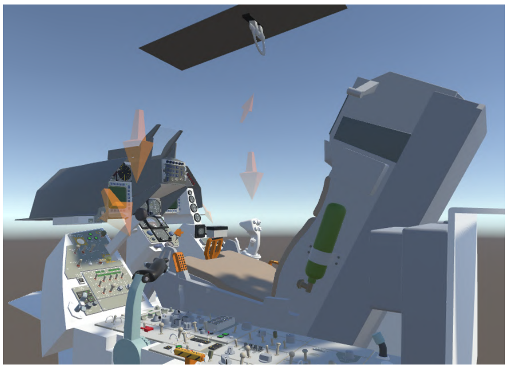

Fig. 1. Side view of the XR-Cockpit. Arrows show the interactive tools included.

camera and a VIVE headset) can provide larger FOV but show the real world with discretization artifacts and low spatial resolution.

Third, we try to verify if, using the two MR headset types, users are still able to complete interactions on real objects.

For this purpose, we designed an evaluation study where subjects had to complete a sequence of manipulations of virtual controllers of the XR-Cockpit in three different experimental conditions: (1) wearing a VR headset displaying VR visualization (2) wearing the same VR headset augmented with co-registered real-world video (video see-through MR), (3) wearing an optical see-through MR headset overlaying the virtual cockpit to the real scene. In all the conditions, the interaction control is implemented via Leap Motion based finger tracking. In the two MR visualization conditions, we also tested how well users can manipulate real objects if a really "mixed" interaction is required, asking them to complete a sequence of actions on tangible objects.

This work presents, therefore, two main contributions for VR/MR research:

- it demonstrates effective solutions for the realization of effective simulation of manipulation based interactive interface
- it presents results of a user study that gives useful hints for the selection of optimal VR/MR solutions in practical applications.

While research efforts have been dedicated in the literature to the study of manipulation-based interfaces [11] and to compare optical vs video see through MR solutions [12], issues related to the integration of these techniques and devices in a realistic industrial training scenario have not been previously investigated.

II. XR-Cockpit

A. Virtual environment and interaction

XR-Cockpit (see Figure 1) is a prototype developed within an industrial agreement with [omitted to maintain anonymity] to demonstrate virtual training applications. Several elements of the cockpit 3D model are interactive and can be manipulated similarly to real ones, exploiting specialized metaphors and

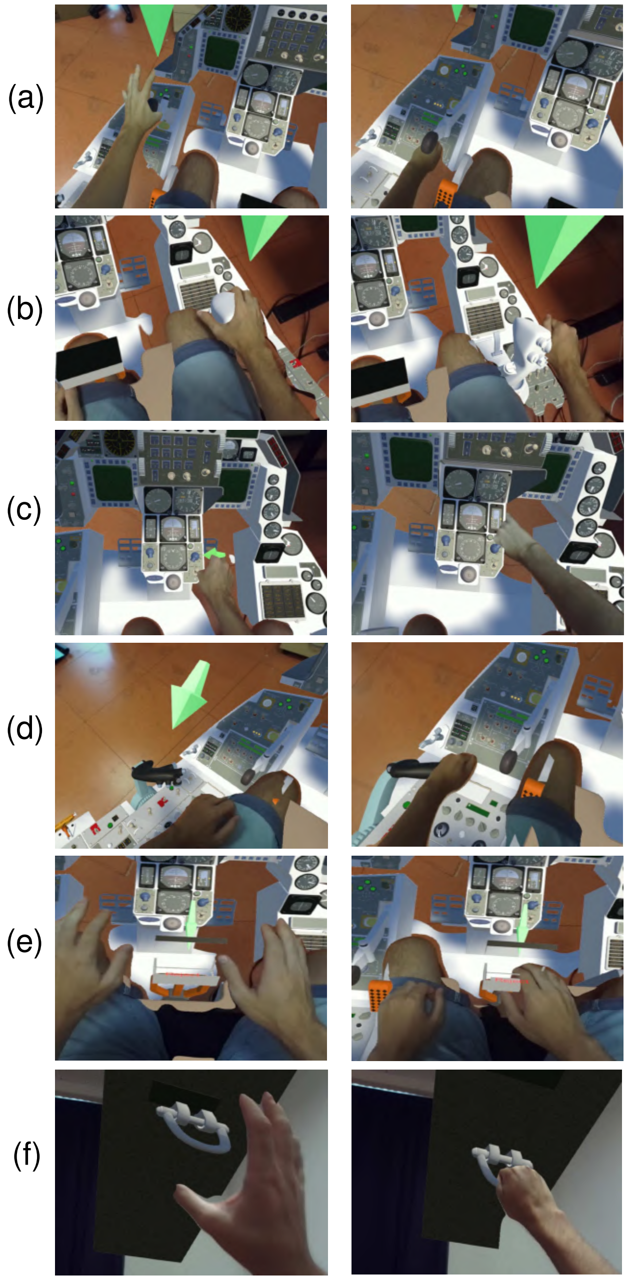

Fig. 2. Virtual manipulations of the six interactive virtual tools as seen by users in video see through Mixed Reality. (a) Left lever. (b) Joystick. (c) Knob. (d) Bottom lever. (e) Box with button inside. (f) Handle.

feedback solutions to overcome the lack of physical contact and haptic feedback.

The interactive tools, that can be used in the simulation mimicking real gestures moving them in the desired position/orientation are the following:

- a lever in front of the seat, in the left part of the cockpit
- a joystick placed on the right of the pilot,
- a knob, in front of the seat,
- a lever situated on the bottom left,
- a box, in front of the pilot, with a push-button inside it,
- a handle placed over the pilot's head that can be dragged

forward and backward.

Figure 2 shows the user view of the tools seen in MR during the interactions. All the scenes and the virtual interactions have been implemented using Unity [9], exploiting the real-time hand tracking provided by a Leap Motion controller, placed on the HMD.

For the interaction with levers, joystick and box, we exploit the Unity physics engine. These objects are moved with direct interaction with the user's hand collision volume, rotating around the intended anchor point on the cockpit employing Unity joint components. For the interaction with the knob, the tracking accuracy was not accurate enough to drive the physics engine, resulting in unstable interactions between the fingertips and knob collision volumes. We decided, therefore, to exploit a pinch gesture recognition and the estimation of the fingers' proximity to the knob at the detection time to enable the action. These data are obtained through the Leap Motion Controller API configured with a proximity threshold of $1cm$. Once the action is enabled, the change in rotation of the hand around the knob rotation axis is used to rotate the knob itself. We adopt a similar solution for the action on the handle. In this case, we exploit the recognition of a grab gesture coupled with an estimation of grab position proximity to the handle to enable the action. We then move the handle according to the hand displacement along the translation axis of the handle. To reduce the impact of missing tactile feedback, we provided the user with auditory feedback consisting of specific sounds associated with the touch and movement of the cockpit controls.

B. Headsets' configurations and visualization modes

In our experiment, we used two different headsets: HTC's Vive [13] and Microsoft's Hololens [14]. Using these described setups, we defined three visualization modes for the XR-Cockpit prototype application, that can be exploited in practical scenarios.

First of all, using the Vive HMD, we created a pure VR visualization mode, displaying only the virtual cockpit immersed in a standard background with an infinite gray floor and a virtual sky (Figure 3), left. In this setup, the position and orientation of the headset in the virtual scene are controlled through the SteamVR Tracking system. We integrated a Leap Motion Controller on the headset, pointing approximately along with the user's gaze (Figure 4 left). With this positioning the user was able to cover all the virtual cockpit interactive area by moving the nominal 150^o Leap Motion FOV towards the points of interest. The hands skeletons are tracked reliably in this area and can be integrated in the virtual scene, controlling the gestural interaction and providing visual feedback. As the Leap Motion sensor is rigidly attached to the headset, its frame can be adjusted in the SteamVR reference system with a simple offset adjustment. In our Vive VR mode, the user sees visual feedback of the hands' position and of the performed actions. We opted for a simple hand skeleton, more realistic representation could be adopted as well, but did not appear more effective in our tests.

Using the same headset, we also created a camera see-through visualization, enhancing the HMD with the video feed of the real-world surrounds using a Zed Mini stereo camera [15]. The resulting video see-through solution has a $85^o \times 55^o$ FOV, with a resolution of 1080×720 pixels. In this Vive-MR mode, the position of the XR-cockpit in the real world is also controlled using the SteamVR tracking system by means of a Vive Tracker device. We relied on the Zed Mini API for depth sorting and blending of rendered and real scene. The API estimates on the fly a 3D model of the use it to determine if the virtual object is occluded by the real one or vice-versa. The reference frame of the Zed Mini reconstruction has been aligned with the Leap Motion/SteamVR references with a manual calibration procedure. This was based on the estimation via offset adjustment the displacement of real and virtual fingers of different users in different positions in the camera frame and the subsequent estimation of an average translation vector to align the two frames. In Vive MR mode we didn't render the hands skeleton captured by the Leap Motion in order to provide a more realistic MR experience.

A second MR visualization has been realized using Microsoft Hololens, an optical see through solution. Thanks to the transparent lenses, it allows direct perception of the environment, superimposing a virtual scene on a smaller FOV ($30^o \times 17.5^o$). We mounted a Leap Motion sensor on it as well (Figure 4 right) to capture the hands and we similarly aligned the Leap Motion and Hololens tracking spaces with an initial calibration procedure. The Hololens setup relies on proprietary Hardware and software for the alignment of the virtual content with the real scene. Similarly to the Zed Mini Api, Hololens API recovers on the fly the 3D structure of the scene and can anchor the virtual content to it. The depth sorting and virtual hands/scene occlusion handling are similar to that of the Zed Mini. Users did not report relevant issues related to the scene blending for both the solutions. We cannot provide snapshots of what is perceived by the eyes in the Hololens MR mode, the superimposed scene appears similar to the Vive-MR one, with the advantage of no resolution issues on the real scene, but with reduced FOV of the virtual one (window effect).

III. USER STUDY

Using the XR-Cockpit, we performed a user study aimed at evaluating how the headset choices and the visualization modes affect the usability of the command tool manipulations. For this purpose, we designed a simple task to be performed on the XR-Cockpit, consisting in completing a sequence of action on the six interactive tools: move the lever of Fig.2 (a) all the way, turn the joystick (b) first on the right and then forward all the way, turn the knob (c) until a specific position is reached, move the bottom lever (d) forward all the way, open the box (e) and bush the button, grab the handle (f) and move forward all the way. To overcome the lack of haptic feedback we added a specific sound related to grab and motion of the tools and another one activated by action completion.

To help the user find the controls quickly and perform the correct sequence we used arrow indicators, displayed near

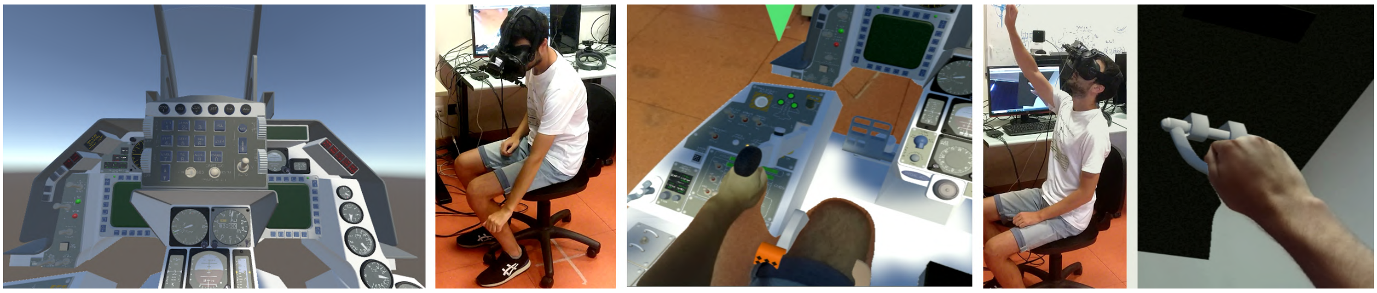

Fig. 3. From left: user view of the XR-Cockpit as seen in VR mode; external views of an user performing two tasks on the cockpit and the corresponding user views in MR mode (cropped)

Fig. 4. The two headsets used fo our tests. Left: HTC Vive with Leap Motion and Zed Mini camera. Right: Microsoft Hololens with Leap Motion.

the targeted control. The indicators are activated only for the currently targeted control and are hidden right after the control is enabled. This ensures that the experiment focuses on the interaction itself rather than learning the cockpit controls layout.

In mixed reality applications, users often interact with both virtual and real objects. The different see-through solutions of our MR configurations, video vs optical, may have an impact on the performances in virtual- and real-world manipulation. For this reason, we included a few additional tasks, namely the manipulations of real-world objects, to be performed only in the MR configurations after the completion of the sequence of manipulation of virtual tools. In this manner, we could verify and compare the performances of the two MR setups in both virtual and real manipulations. In MR configurations the virtual cockpit appears to the user with a real-world table on the right side (see Figure 5). In MR configurations. After the completion of the previously described sequence, we asked subjects to look to the right and: (g) click a button on a mouse, (h) stack three small boxes (Figure 5 b), (i) open the capsule compartment on a coffee machine (Figure 5 c), (l) put a glass in the right place, (m) close the compartment and (n) click the mouse button again (Figure 5 d).

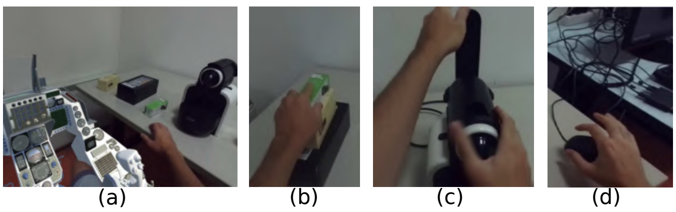

Fig. 5. The real interaction task as seen from the camera see-through headset: (a) the table is seen on the right of the cockpit. After clicking on a mouse button, the subject has to stack three small boxes (b) and to perform some actions on a coffee machine (c) and then click again on the mouse button (d).

The idea of our study is to verify how users can perform the sequence of actions on the virtual tools with the three different hardware and visualization modes and to check if the MR setups can be used effectively for real world manipulation as well. We aim to verify the following hypotheses:

H1 there is no difference in the virtual interaction tasks completion times between the Vive VR and Vive MR configurations; in principle, the current HMD technology should limit problems in VR depth perception [16], and the presence of complex virtual objects in our scene should provide a good spatial context in VR;

H2 subjects in Vive MR conditions are faster than users in Hololens MR conditions due to the increased FOV helping to find target objects;

H3 subjects in Hololens MR conditions are faster than users in Vive MR conditions for the manipulation task with tangible objects, due to the better spatial perception without camera distortion.

IV. Experimental Design

We selected 24 subjects, with age from 21 to 45, of which 20 males and 4 females. 11 subjects had previous experience of using HMD devices for gaming or research, 13 had no experience of VR/MR applications. Each subject performed the virtual manipulation tasks three times for each of the three different conditions: VR Vive (VR-V), MR Vive (MR-V), MR Hololens (MR-H). We recorded completion times of the complete sequence and all the tasks described in Section III. Before each experimental condition, subjects had 2 minutes of training time to understand how to perform the expected actions on a scene with interactive devices that could be manipulated similarly to those included in the cockpit scene. We made subjects aware of the Leap tracking volume being related to their gaze direction and during the experiment we recorded any loss of hand tracking due to this issue for each subtask. The order in which the users had to test the different conditions (VR-V, MR-V, MR-H) was decided for each subject following a Latin square design to fully counterbalance possible order biases. After completing the virtual manipulation task, in the two MR conditions, subjects had to perform the manipulation task on the real objects on the right side of the virtual cockpit.

After the completion of all the tasks in the different conditions, the subjects were asked to fill a questionnaire to record their feedback. We use scores on a Likert scale from 1 (poor) to 5 (excellent) on the following aspects:

- ergonomic comfort of the HMDs (Vive vs Hololens);
- ease of execution of the virtual manipulation tasks in the three conditions;
- visual comfort during the execution of the virtual manipulation tasks in the three conditions (visibility of the objects in the scene, clearness of sight, easiness of visual perception);
- easiness of execution of the tasks on real objects (MR conditions only);
- visual comfort during the execution of the real manipulation (MR conditions only).

The last four questions asked to indicate the preferred setup for the virtual task, the preferred setup for the real task, the overall preferred MR setup and the preferred condition (MR vs VR) for the completion of the virtual tasks on the cockpit with the Vive HMD. We also collected open comments from the subjects and generic opinions on the potential applications of the system.

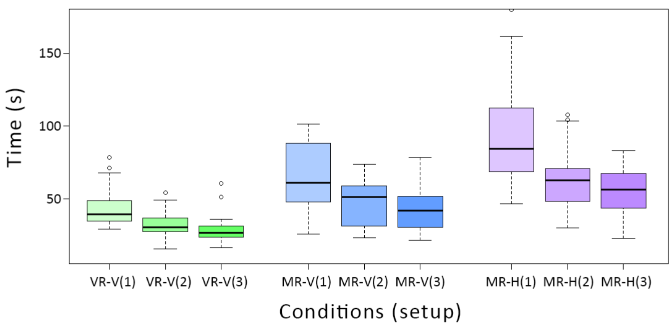

Fig. 6. Box plots representing the distributions of the execution times of the virtual interaction in the three different conditions, VR with Vive (VR-V), MR with Vive and Zed Mini (MR-V) and MR with Hololens (MR-H), for the three task iterations labeled (1),(2) and (3).

V. Results

The system was well accepted by users in all display configurations. It is worth noting that easiness of task completion was scored sufficient in all configurations and we did not encounter relevant issues as users not able to complete the tasks in a limited time. This means that the virtual manipulation metaphors were easily understood and that the visual feedback was of sufficient quality for our application. The hand tracking worked well in all configurations. We recorded only 0.02% of actions suffering delays caused by the failure of hand tracking. This typically happens when the hands leave the interaction area. The quantitative analysis of the user study is reported in the following subsections and discussed in Section VI.

A. Execution times

The average and median execution times for all tasks in all configurations, along with quartiles of the related distributions are reported are represented as box plots in Figure 6, that show all the relevant differences. The main findings are that the average execution times in the VR-V condition are sensitively lower than those obtained in the MR-V condition, while the average times in the MR-V condition are sensitively lower than those obtained in the MR-H runs.

In our experiment, the same tasks were repeated three times in each experimental condition. After the first repetition, we measured a decrease in completion times likely due to the knowledge of the tasks and the objects' locations. This effect is present, but significantly less, in the third repetition.

We carried out a statistical analysis on the completion times to identify significant differences between them. We compared the average times obtained in the different conditions at corresponding task iterations to verify our hypotheses H1, H2 and H3. We also compared the average times obtained for each fixed condition, in the three iterations, in order to capture learning effects. As the data didn't satisfy the normality hypothesis, verified with a Shapiro-Wilk test, we estimated the statistical significance of the differences with a non-parametric test, the Wilcoxon signed-rank. All the p-values reported next were corrected following the Bonferroni-Holm post-hoc test in case of multiple comparisons.

First, we compared the average times for the same iteration, but in different conditions, and reached the following conclusions:

- VR-V, on average, has significantly lower execution times compared to MR-V in all the three iterations #1,#2 and #3 with p-values equals to 0.0027, 0.0011 and 0.0032, respectively;
- VR-V, on average, has significantly lower execution times compared to MR-H in all the three iterations #1, #2 and #3, all with a p-value of 0 approximated to the 4th digit;
- MR-V, on average, has significantly lower execution times compared to MR-H for all the iterations #1, #2 and #3, with p-value equals to 0.0054, 0.0067 and 0.0134, respectively.

These results confirm H2 and H3 while H1 is not supported by the test outcomes. We will discuss this in Section VI.

Then, we compared the execution times obtained with the same condition across the three iterations, and we reach the following results:

- for each condition VR-V, MR-V, MR-H, the first run has a significant higher execution time compared to the second, respectively with a p-value of 0.0003, 0.0226 and 0.0007 all approximated to the fourth digit;
- for each condition VR-V, MR-V, MR-H, the second run didn't have any significant difference compared to the third, with all p-value > 0.05.

These results show that our hypothesis H1 is not true. Whole task completion is faster in VR-V conditions. We discuss this in Section VI. H2 and H3 are verified. To evaluate the effectiveness of the different interactive elements of the

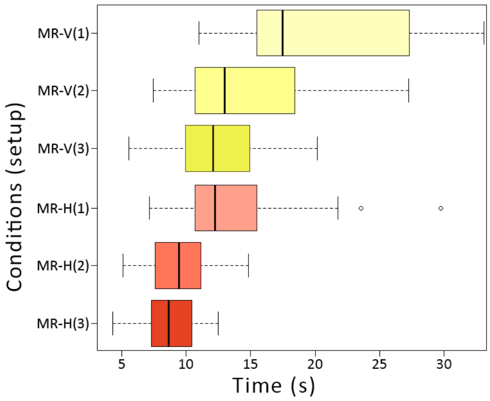

Fig. 7. Box plots representing the distributions of the execution times of the task on real objects in the two different conditions, MR with Vive and Zed Mini (MR-V) and MR with Hololens (MR-H), for the three iterations labeled (1),(2) and (3).

XR-Cockpit in the three visualization conditions, we look at the performances in single tasks. Looking at single averages, shown in Figure 8, and performing Wilcoxon signed-rank tests to compare them, we see that the differences of the average times in the single subtasks are similar to those of the complete repetition, with a single exception: in task (c), the knob rotation, there is no significant difference between MR-H and MR-V.

We analyzed the data for real-world interaction in the same manner we did for the virtual ones. Results are visualized with box plots in Figure 7. For these times we used the Wilcoxon signed-rank test since data did not satisfy the normality hypothesis. The test revealed that, on average, the execution time is lower for the MR-H condition compared to the MR-V with a p-value of 0.0005, 0 and 0.0005 for the three repetitions. Comparing differences across repetitions, we found a significant variation between the first and second run for both the MR-V and MR-H conditions, respectively with p-values of 0.006 and 0.0007, approximated to the fourth digit, and no difference between the second and the third run, with a p-value > 0.05 in both cases.

B. Questionnaires' data

Figure 9 summarizes the questionnaire ratings given on a 1-5 Likert scale The differences between corresponding Likert-scale ratings have been analyzed with the Generalized Linear Model (GLM) statistic test. Scores are high for all the experimental conditions, consistently with the free comments collected from users that indicate good system usability in all the three conditions. We did not record significant differences in the ergonomic comfort of the HMD. For the virtual task (see Figure 9), significant differences were found between the easiness of the task and the visual comfort of the three different experimental conditions. Users rated higher the easiness of execution and the visual comfort for the VR setup against both the MR conditions. A significant difference was found for the visual comfort of MR-V condition rated higher than the MR-H. The easiness perceived in the MR-V condition is also higher than the one reported for MR-H, but the difference was not statistically significant.

For real-world manipulations, the MR-H setup was preferred to MR-V for both easiness of execution and visual comfort (see Figure 9). In these tasks, the ratings follow the same trends as execution times: scores are higher when the execution times are lower.

Finally, we analyzed the answers to the multiple-choice questions about the various setups. For the virtual manipulations, users clearly preferred the VR-V condition (23 votes) to both the MR-V (0 votes) and MR-H conditions (1 votes). For real-world manipulations, users prefer MR-H over MR-V, with 20 and 4 votes respectively. In both cases a Pearson chi-squared test demonstrate a statistical significance of the results. User has no significant preference when ranking their overall preferred MR setup and the overall preferred mode between VR and MR, with both p-value above 0.05. However, a clear majority of users preferred the VR interaction (67%) and the Vive/Zed Mini setup (58%) as MR viewer.

A potentially relevant result is that we did not find statistically significant differences in performance between the group of experienced and non-experienced users ($p < 0.05$). This may be due to the fact that the interactive manipulation

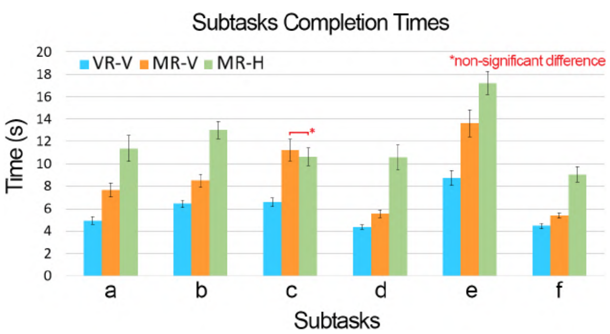

Fig. 8. The relative efficiency of the manipulation in the three conditions is almost always consistent in the six subtasks. The only exception is the knob rotation, where the interaction speed in MR-H and MR-V is not different. One possible reason is related to the fact that the interaction is performed in a limited and central region, reducing the issues related to the small FOV of the Hololens.

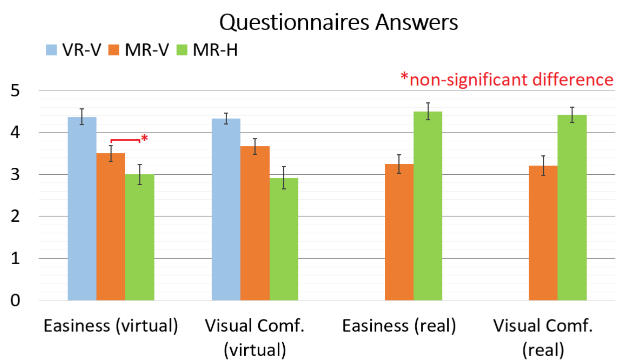

Fig. 9. In the "virtual" task we recorded significant differences among the subjective scores evaluating visual comfort in the three different conditions (VR-V better than MR-V better than MR-H) and a significantly higher ease of use in the VR-V condition.

involved in these tasks is not common in gaming or typical VR/AR experiences.

VI. Discussion

A. Overall system usability

In free comments, subjects stressed on the fact that the interactive experience, despite a few problems reported, was considered ready to be employed in commercial applications. This is consistent with the positive scores given to ease of use and visual comfort for all experimental conditions. The key factors enabling good usability are the quality of Leap Motion finger tracking, and the replacement, in the most challenging cases, of the simulated physical interaction with gesture recognition and subsequent motion mapping. This makes the interaction reasonable even in case of tracking inaccuracy. The use of visual and auditory feedback was considered useful by users in free comments.

B. Outcomes of the statistical tests and hypotheses verification

We compared different MR/VR setups to complete tasks that could be implemented in real training stations. The use of the same hand tracking configuration and the same software solutions ensured that the differences recorded for the different experimental conditions only depended on the visualization choices. However, it is not always obvious to determine the reasons for the differences recorded.

Considering our hypothesis H1 (no difference between Vive VR and Vive MR), we found that it is refuted, as the VR solution was sensitively better both in performance and user preference. This seems in contrast with the results of previous studies, e.g. [10], that show an advantage in the use of MR visualization for docking tasks, however, it should be noted that both the tasks and the visualization solution in [10] are quite different from ours. Accurate docking is surely more influenced by the correct spatial perception than our manipulation tool-based interactions and the VR background used in [10] is more cluttered.

A possible reason for the superiority of the VR interaction in our test is related to the different visual feedback for the hands. Hands' appearance has a relevant effect on interactive task performances in VR as shown in [17]. In our VR configuration, the skeletons of the captured arms and hands are rendered, while in the MR configurations the real hands are seen (see Figure 2). While the registration accuracy of the two hands is sufficient to have a good sense of presence and an effective manipulation, it is clear that in MR there may have been inaccuracies leading to the necessity of more careful adjustments of the hand position to obtain the correct interaction. Inaccurate registration may also lead to the incorrect depth-based fusion of real and synthetic objects, even if the subjects did not report related issues.

Hypothesis H2 (Vive MR better than Hololens MR in our task), was clearly confirmed by our tests. The reason is probably related to the narrow FOV of the Hololens. This is also supported by the task analysis, as the difference is not significant for the action on the knob that is directly in front of the user and easier to find with the Hololens. However, other factors may result in different performances.

H3 (Hololens MR faster for tangible objects' manipulation), was also confirmed by the experiments. The distorted spatial perception with the see-through solution requires the user to rely on visual feedback only and to deal with the visual-proprioceptive discrepancy. Some users reported a bad feeling caused by the touch of real objects when seen on the see-through display, presumably due to a mismatch between visual and tactile feedback.

It may be interesting to note that, while the efficiency of the interaction in the three conditions decreases with task iterations, the differences across the conditions is unchanged. The fact that the differences in execution times between the second and the third iterations are not significant may indicate that the difference may remain after training. This was not obvious, as one could expect that, if the advantage of the increased FOV of MR-V against MR-H is only related in the easier search of the tools to be used in the 3D space, the advantage should disappear after one or two iterations, when the user knows, even approximately, the objects' position. As reported by some users, the problem is possibly related to the narrow FOV of the first generation of Hololens. The FOV in question is, according to some users, so narrow that it makes difficult not only to locate the object but also to perform the actual manipulation action, as the context provided to put the hand in the correct position is not sufficient. This fact is supported also by subtask (c) results of the virtual task, that, as reported in V-A, is the only one not displaying any significant difference in performance across the three setup. We think this is due to the subtask being placed in front of the user and, therefore, easier to locate. Moreover, we use gesture recognition since the physical simulator was not accurate enough, which means that an accurate visual superimposition of the real hand and the virtual tool is not required.

C. Limitations

The results of our study depend on the chosen devices. Improvements in tracking or display quality may change the relative scores of video and optical see-through. As no significant concerns on the spatial resolution of the displays have been reported, the use of VR HMDs, newer than the Vive, should not have changed the results relevantly, while we expect that the use of MR HMDs with increased FOV, like Hololens 2 or Magic Leap, could make optical see-through visualization more acceptable, even if the FOV of these solutions is still limited.

Another factor potentially affecting the results is the latency of the devices. While both Hololens and Vive motion-to-photon latency, according to our tests, is almost negligible (less than 2ms and 7ms on low load), it is not the case for the video see-through latency of our MR setup using Zed Mini. The Zed Mini has a nominal motion-to-photon latency of 60ms, that is low compared to other available solutions, but that might explain the problems with tangible manipulations in the video see-through conditions.

The registration of the virtual and real scene was not adapted to the single users and this may have created issues in particular cases. We recorded in free comments some issues related to unexpected gesture outcomes due to inaccurate location of the hand respect to the pivot point of the manipulated tool.

D. Design guidelines for VR/AR training stations

Our test application is a prototype of a potential training station that can be used to learn complex procedures on real interfaces. The implementation of this kind of tools on low-cost off-the-shelf hardware could enable a large diffusion of them in different contexts. From the outcomes of our tests, we derived some guidelines/suggestions that could be useful for designers of such systems.

Smart manipulation solutions can solve usability issues while keeping sufficiently natural interaction When the use of finger tracking and physics simulation fails in providing a smooth interaction, exploiting gesture recognition and motion remapping can result in reasonable user experience. Auditory feedback may also help with overcoming the lack of haptic feedback in the virtual interaction.

VR is the right choice when MR is not necessary Our experiments confirm that there are no relevant advantages in adding real context when acting only on virtual objects, while the poor registration of virtual hands with the real ones seen in MR can create problems in MR setups. Users reported no interaction issues related to the quality of the rendering and the interaction with the virtual devices was both faster and preferred in VR.

Superimposition of tracked hands to real ones in AR manipulation may be useful as feedback for users. Poor registration could be the cause of worse performances obtained in the Vive AR setting vs Vive VR settings. If the application requires virtual object manipulation immersed in a real-world context, a solution can be to superimpose the virtual hand over the real one in MR to provide more feedback. We plan to check quantitatively this effect in a future user test.

A wide enough FOV is highly recommended for MR applications Despite the high quality of the Hololens tracking and rendering, the limited FOV makes the MR interaction quite difficult and problems seem to persist after a few iterations of the task. The video see-through setup tested seems sufficiently flexible and practical for MR applications and there are not cheap optical solutions currently on the market offering a FOV comparable to the Zed Mini. A proper test evaluating the effects of the different FOVs of the devices available for the MR visualization on the specific tasks of interest could help designers to select the right solution for each application.

Video see-through MR works well but presents relevant issues when interactive task requires the manipulation of real objects. In the case of MR tasks with a majority of real manipulations, an optical see-through solution is a better choice.

VII. Conclusions

We presented XR-Cockpit, an immersive MR/VR training station prototype with interactive manipulation of command tools and the report on a specific user test aimed at evaluating the effects of different visualization options on the interactions designed for our prototype. We believe that the results of our study can be useful to navigate the differences in setups when designing applications involving complex interactions. Our work points out some key pros and cons of the different setups, identifies potential reasons for the differences in their performances and also indicates directions for future work.

Acknowledgements This work has been partially supported by The Edge Company and by the project MIUR Excellence Departments 2018-2022.

References

[1] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi, “A systematic review of augmented reality applications in maintenance,” *Robotics and Computer-Integrated Manufacturing*, vol. 49, pp. 215–228, 2018.

[2] M. Juraschek, L. Büth, G. Posselt, and C. Herrmann, “Mixed reality in learning factories,” *Procedia Manufacturing*, vol. 23, pp. 153–158, 2018.

[3] P. Fraga-Lamas, T. M. Fernández-Caramés, Ó. Blanco-Novoa, and M. A. Vilar-Montesinos, “A review on industrial augmented reality systems for the industry 4.0 shipyard,” *IEEE Access*, vol. 6, pp. 13 358–13 375, 2018.

[4] E. Barsom, M. Graafland, and M. Schijven, “Systematic review on the effectiveness of augmented reality applications in medical training,” *Surgical endoscopy*, vol. 30, no. 10, pp. 4174–4183, 2016.

[5] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, “The effectiveness of virtual and augmented reality in health sciences and medical anatomy,” *Anatomical sciences education*, vol. 10, no. 6, pp. 549–559, 2017.

[6] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, “A survey of augmented, virtual, and mixed reality for cultural heritage,” *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 11, no. 2, p. 7, 2018.

[7] L.-H. Lee and P. Hui, “Interaction methods for smart glasses: A survey,” *IEEE Access*, vol. 6, pp. 28 712–28 732, 2018.

[8] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster, “A dose of reality: Overcoming usability challenges in vr head-mounted displays,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 2143–2152.

[9] “Unity web site,” https://unity.com/, accessed: 2019-08-30.

[10] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato, “Augmented reality versus virtual reality for 3d object manipulation,” *IEEE transactions on visualization and computer graphics*, vol. 24, no. 2, pp. 1038–1048, 2017.

[11] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge, “A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments,” in *Computer Graphics Forum*, vol. 38, no. 1. Wiley Online Library, 2019, pp. 21–45.

[12] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva, “Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 1, pp. 125–139, 2013.

[13] P. Dempsey, “The teardown: Htc vive vr headset,” *Engineering & Technology*, vol. 11, no. 7-8, pp. 80–81, 2016.

[14] B. C. Kress and W. J. Cummings, “11-1: Invited paper: Towards the ultimate mixed reality experience: Hololens display architecture choices,” in *SID symposium digest of technical papers*, vol. 48, no. 1. Wiley Online Library, 2017, pp. 127–131.

[15] “Stereolab website,” http://www.stereolabs.com, accessed: 2018-08-11.

[16] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger, “Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr,” in *IEEE Conference on Virtual Reality and 3D UserInterfaces*, 2019.

[17] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han, “Effects of virtual arm representations on interaction in virtual environments,” in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 2017, p. 40.