Check for updates

An empirical evaluation of two natural hand interaction systems in augmented reality

Ramiro Serrano[1] · Pedro Morillo[2] · Sergio Casas[2] · Carolina Cruz-Neira[3]

Abstract

Human-computer interaction based on hand gesture tracking is not uncommon in Augmented Reality. In fact, the most recent optical Augmented Reality devices include this type of natural interaction. However, due to hardware and system limitations, these devices, more often than not, settle for *semi-natural* interaction techniques, which may not always be appropriate for some of the tasks needed in Augmented Reality applications. For this reason, we compare two different optical Augmented Reality setups equipped with hand tracking. The first one is based on a Microsoft HoloLens (released in 2016) and the other one is based on a Magic Leap One (released more than two years later). Both devices offer similar solutions for the visualization and registration problems but differ in the hand tracking approach, since the former uses a *metaphoric* hand-gesture tracking and the latter relies on an *isomorphic* approach. We raise seven research questions regarding these two setups, which we answer after performing two task-based experiments using virtual elements, of different sizes, that are moved using natural hand interaction. The questions deal with the accuracy and performance achieved with these setups and also with user preference, recommendation and perceived usefulness. For this purpose, we collect both subjective and objective data about the completion of these tasks. Our initial hypothesis was that there would be differences, in favor of the isomorphic and newer setup, in the use of hand interaction. However, the results surprisingly show that there are very small objective differences between these setups, and the isomorphic approach is not significantly better in terms of accuracy and mistakes, although it allows a faster completion of one of the tasks. In addition, no remarkable statistically significant differences can be found between the two setups in the subjective datasets gathered through a specific questionnaire. We also analyze the opinions of the

---

✉ Sergio Casas
  Sergio.Casas@uv.es

1 Emerging Analytics Center (EAC), University of Arkansas (UALR), Little Rock, AR, USA

2 Department of Computer Science, University of Valencia (UV), Valencia, Spain

3 Agere Chair at the Department of Computer Science, University of Central Florida (UCF), Orlando, FL, USA

🖄 Springer

participants in terms of usefulness, preference and recommendation. The results show that, although the Magic Leap-based system gets more support, the differences are not statistically significant.

Keywords Augmented reality · Natural interaction · Gesture tracking · HoloLens · Magic leap

1 Introduction

Augmented Reality (AR), as defined by Azuma [7], is a computer-based technology that simultaneously combines real and virtual objects which are registered in 3D and are interactive in real-time. This widely accepted three-feature definition is directly related to the three major technical challenges that AR faces. The first one is the meaningful combination of virtual and real objects, which is called the *visualization problem*. A solution to this problem requires to build display devices that are able to draw virtual objects over a view of the real world (or onto the real world itself as in spatial AR [9]). In order to accomplish this, a realistic and seamless combination of virtual and real objects is necessary. The second challenge is the registration of virtual objects within a real 3D world, which is called the *registration problem*. This problem requires tracking the user movements so that the application can identify where the objects of the real world are located with respect to the user. The last challenge is to achieve an interactive application. This is closely related to the *interaction problem*, which implies that the user should - ideally - be able to perform a seamless interaction with both real and virtual objects. Interfaces are said to have *seams* if there is a functional, temporal, spatial or cognitive element that forces the user to change the way in which he/she interacts with the interface [23].

None of these problems is easy to solve. The visualization problem, for instance, includes also other problems, such as the occlusion problem [19, 59]. There are several technical solutions to perform this integration of virtual and real objects. These solutions range from camera-based solutions (as in mobile AR [13]), light projection (as in spatial AR [29]) or optical solutions based on half-silvered mirrors [46] or on see-trough Head-Mounted Displays (HMDs) [18]. It is also possible to use a non-see-through HMD to create an AR application [65].

The registration and tracking problem has been a major problem in AR research over the last twenty years. Although outdoor tracking is still an issue [31], there are currently good solutions for indoor tracking that are quite reliable [32].

With respect to human interaction in AR systems, many possible interaction systems can be considered *natural*, such as voice recognition, eye/gaze tracking, body movement, facial expression and, of course, hand and gesture tracking. In this area of multimodal interaction, there are still many unsolved problems, such as tactile feedback, device calibration or decoupling between real and virtual actions. For these reasons, the goal of achieving a seamless natural interaction between human and virtual objects is still unattainable. However, interaction systems based on hand and gesture tracking are not uncommon in AR environments.

Natural hand interaction systems can be classified in *metaphoric* and *isomorphic* [35]. A metaphoric hand interaction system is defined as one that "bases input actions on image schemas - mental models formed from repeated patterns in everyday experiences - and system effects on related conceptual metaphors". In this regard, as HoloLens only detects a limited set of gestures representing common actions, it can be considered a metaphoric system. On the

contrary, *isomorphic* interaction systems can be defined as those that perform "one-to-one literal spatial relations between input actions and resulting system effects". Meta 2 and Magic Leap One could fall in this category.

Understanding the benefits of using either of these two types of interaction techniques is key for developing successful AR applications. For this reason, we perform in this research a comparison of these two natural hand interaction systems. The analysis is performed using two AR devices - from two different manufacturers -, a Microsoft HoloLens (released in 2016) and a Magic Leap One (released in 2018), that offer similar solutions for the visualization (by means of an optical see-through HMD) and registration problems (using Simultaneous Localization and Mapping (SLAM) algorithms). These two devices are representative of the two aforementioned interaction techniques. The former uses a metaphoric hand-gesture tracking system and the latter relies on an isomorphic approach. In addition, although this is not the goal of the experiment, quantifying the amount of improvement in natural hand interaction that the newer AR device (the Magic Leap One) achieves with respect to the older (the Microsoft HoloLens) could serve as an indirect indicator of the maturity of this technology, since gesture tracking is one of the features in which both systems present more differences.

The rest of the paper is organized as follows. Section 2 reviews related works using HoloLens and Magic Leap and also reviews studies on AR interaction, with special focus on hand interaction. Section 3 describes the materials and methods utilized to develop the tools employed in the experiments. Section 4 details the experimental study. In section 5, the results of these experiments are presented and discussed. Finally, section 6 outlines the future work and shows the conclusions of the paper.

2 Related work

Several applications and research works have been published using a HoloLens. A sizeable amount of this research is concentrated in the medical field, in which many applications have been recently proposed using this optical AR device [1, 2, 14, 20, 26, 30, 36, 47, 50, 67]. The fact that this device allows physicians to experience virtual content without having their hands busy holding hardware, makes it very convenient for medical applications, and explains why is so widely studied in the medical field. There exist, however, many other applications of this optical AR system [5, 6, 22, 41, 55, 64, 66].

In addition, some research works have analyzed the performance of this optical AR device. For instance, in [33] a comprehensive technical evaluation of HoloLens's performance is reported. The authors of this work analyze head tracking, environment reconstruction, spatial mapping and speech recognition. They find that head tracking (using OptiTrack Flex V100 [42] as ground truth) is more properly performed at low head-movement speeds, that the 3D reconstruction of the environment is most precisely for flat surfaces under bright conditions, and that speech commands have correctness rates of 74.47% and 66.87% for user-defined and system-defined commands, respectively. Spatial mapping and tracking accuracy are also evaluated in [16] for an AR assembly application. They find that the HoloLens system was not accurate enough to create a mesh of some intricate parts and, thus, a Vuforia marker-based tracking plug-in was necessary. Therefore, they conclude that the device is not yet ready for deployment in a factory assembly setting where high precision is necessary. A different analysis is performed in [11]. This work focuses on gesture tracking and in bimanual gestures

in particular. The authors implement and evaluate five techniques for rotation and scale manipulation gestures on the Microsoft device, conducting a study with 48 users. They find that hand-tracking losses occurs frequently in two-handed gestures due to the limited field of view (FOV) of the device. In fact, the manufacturer discourages the development of two-hand gestures [39]. Nevertheless, this work also reports that certain bimanual techniques "perform comparatively to the one-handed baseline technique (the one-handed *wireframe cube* technique currently in use on standard HoloLens applications shipped with the device) in terms of accuracy and time".

Similar problems are found in a recent paper comparing a Microsoft HoloLens and a Virtual Reality (VR) Table [54]. This study analyses the responses of 82 participants using virtual cadavers for anatomy training in both the AR setup (with HoloLens) and the VR setup. Some of these participants commented about the limited FOV (around 30°) of the HoloLens. Some others indicated that they expected a higher accuracy both in head tracking (especially when fast head movements are performed) and in hand gesture recognition. These and other problems caused a preference for the setup based on the VR Table. The results of this previous study encouraged us to perform the work presented here, since the use of the HoloLens hand interaction system in the AR setup seemed to be one of the key points that favored the VR-based setup in [54].

As it can be seen, there is a substantial amount of research involving or using the HoloLens. On the contrary, there are very few academic works using or dealing with the Magic Leap One, since it is a newer device. However, an interesting application can be found in [51] where the Magic Leap is combined with an Apple Watch and a touchless ultrasonic haptic device, in order to allow the user to experience an holographic view of his/her own heart beating. The touchless haptic device combined with the Magic Leap allows an untethered natural hand interaction with some tactile feedback. As in the case of HoloLens, the medical field seems to be especially attractive, such as in [58], where a remote medical monitoring system using Magic Leap is presented, or in [21], an application where an AR system based on Magic Leap is implemented for people with Autism Spectrum Disorder (ASD). Other works using or dealing with this device can be found in [8, 15, 56, 57]. However, to the best of our knowledge, no academic work has yet reported a detailed scientific analysis of the technical features of this device, nor a comparison between a Magic Leap One AR application and a HoloLens-based one. For this reason, we believe our contribution is meaningful.

Regarding interaction in AR/VR systems, it is generally accepted that natural interaction techniques increase the level of fidelity of AR/VR applications. However, there can be cases where non-natural interaction systems could be preferred over semi-natural or natural interfaces [37, 48]. Furthermore, it is hypothesized that interaction fidelity may follow a U-shaped curve [38] where mid-fidelity techniques are the worst and systems with low or high levels are preferred by users, because unlike semi-natural interfaces, both natural and non-natural interfaces feel familiar. This *uncanny valley* [40] appears also in the perception of the aesthetics of robots and has been widely studied.

With respect to natural hand interaction in AR/VR systems, many solutions have been proposed. Two main technologies can be identified in the implementation of hand gesture recognition [61]: those based on wearables such as data gloves [34, 44] and those relying on sensors such as video cameras, depth cameras or infrared sensors [27, 53, 60]. A third category is also possible for those combining both methods ([24, 62]). Glove-based interaction systems were first developed in the 1980s, but today there is a preference towards the use of sensors and cameras for hand tracking, due to freedom of movement. Wearables, on the other hand,

can provide higher accuracy and reliability. In addition, they can provide haptic cues [10], which is an advantage over vision-based systems.

With respect to comparisons in natural interaction for AR systems, some authors have performed comparisons between different interaction systems. For instance [45] compares a free-hand gesture-based interaction technique with a multimodal gesture-speech interaction system. The results show that both systems have strengths and weaknesses. Similarly, [28] compares a multimodal interaction system combining free-hand gesture and speech input with speech-only and 3D-hand-gesture-only interaction conditions (across a series of object manipulation tasks). The results show that the multimodal interface (MMI) is more usable than the gesture-only version. Nevertheless, the MMI was neither more effective nor more efficient than the speech-only system. Another experiment in the same line can be found in [12]. The results show that speech outperforms gesture in terms of accuracy but the simplicity of gestures compensates accuracy with speed. Therefore, they conclude that both gesture and speech are effective interaction modalities for performing simple tasks in AR applications. A somehow similar experiment is performed in [52] using a HoloLens. The authors conclude that users were faster using the hands-free approach (voice control) than using the manual (gestural) interaction.

Other authors focus their research on the hand gesture interaction system. For instance, in [43] two gestural interaction mechanisms are evaluated and compared. However, the experiment is not performed for AR applications and uses a data glove. In a more recent work, a Leap Motion Controller and a smartphone are used to implement and AR application to interact with 3D models from a museum collection [27]. The results show that the system is well accepted by the museum visitors. Another interesting experiment is shown in [4] where the authors analyze freehand grasping in exocentric Mixed Reality, using a Microsoft Kinect. Other authors focus in the learning effects in the use of gesture interaction in AR, such as in [49]; this study shows that users learn how to use hand-based gesture control in a short time.

Finally, a few works are focused specifically on comparing two or more hand interaction devices/mechanisms for virtual environments. Three different hand gesture recognition devices - Leap Motion, Microsoft's Kinect, and Intel's RealSense - are analyzed in [25]. The authors conclude that Leap Motion is the best device in the context of game design. In [3] an experiment is designed to compare two different gesture-based approaches (high and low level of naturalness) in AR using a Leap Motion Controller for three basic actions: translation, rotation and scaling. The authors conclude that participants struggled with highly natural approaches. Therefore, although these approaches were more enjoyable, participants would choose the less natural approaches for future interactions.

A much more rigorous approach to this matter is shown in [17]. This paper is especially interesting because it shares some similarities with our approach. It shows an experimental comparative evaluation between a metaphoric hand interaction system (represented by a Microsoft HoloLens) and an isomorphic hand interaction system (represented by a Meta2 device) for three different types of tasks: move, scale and rotate a virtual cube. The results show that both interaction systems present strengths and weaknesses. The isomorphic paradigm is perceived as more natural and usable (with a higher usability score and a lower task-load index) and is more accurate for the displacement task, whereas the resize task is more accurate under the metaphoric paradigm.

There are several differences between the work shown in [17] and our approach. First, an obvious difference is that we compare different devices. Second, the tasks that we compare in this work are more complex. Finally, the methodology is different, since [17] uses objective

measurements and two generic subjective evaluation systems - the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX) -, whereas we analyze much more objective indicators and use subjective questionnaires designed specifically for AR/VR applications.

3 Materials and methods

As previously mentioned, two different hardware setups are used in this research in order to assess the differences regarding selection and displacement of objects in AR: A HoloLens-based system and a Magic Leap-based system. Both devices use the same optical AR paradigm, use similar head-tracking procedures and will run the same software application in these experiments. These two setups represent two of the most common see-through HMD used in AR applications.

3.1 HoloLens-based system

The first hardware station was setup by means of a table and a Microsoft HoloLens (version 1) device. This device consists of a pair of Mixed Reality smart glasses developed and manufactured by Microsoft. It is a see-through HMD with a built-in PC with Windows 10, a FOV of approximately 30×17 degrees, a display with 1268×720 pixels per eye, a two hand-gesture recognition system with limited hand tracking and a time-of-flight depth sensor. Figure 1 shows a user testing the HoloLens-based station.

3.2 Magic leap-based system

The second hardware station was set up using a table and a Magic Leap One device (see Fig. 2). The Magic Leap One is a three-piece system that includes a headset called *lightwear*, a small wearable computer called the *lightpack*, and a handheld controller. The headset is studded with tracking cameras for mapping the environment, as well as inward-facing eye-

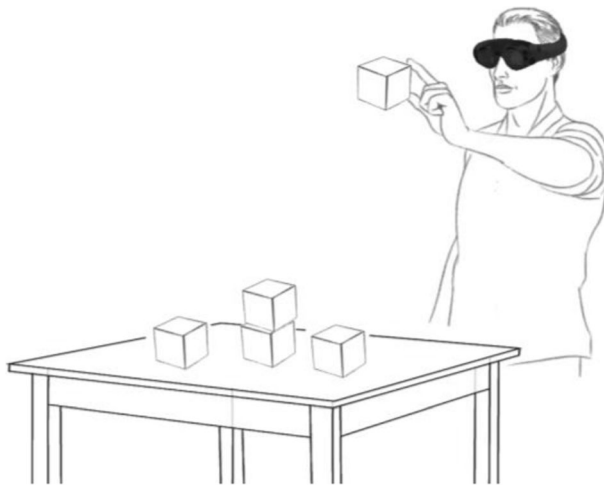Fig. 1 A person using the HoloLens-based application at UALR

Fig. 2 Schema of the Magic Leap-based setup. The schema of the HoloLens-based setup is similar: the AR glasses change, but the rest of the elements (table, software, tasks, etc.) remain the same

tracking cameras. The darkened lenses are inset with small glass waveguides, which Magic Leap manufacturer calls photonics chips. The sensors are located directly on the headset.

The device displays are based on the virtual retinal display (VRD) technology, which draws a raster display directly onto the retina of the eye. Pupil tracking technology is also incorporated in the headset. Thus, it is possible to know where exactly a user is looking at, although this is not used in our experiments.

This device provides a FOV of 40 × 30 degrees, a display with 1280 × 960 pixels per eye, a controller with 6 degrees of freedom (not used in the experiments), a two hand-gesture recognition system including finger tracking (with 3 joints per finger), and a time-of-flight depth sensor. It sells at $2300 and is a serious competitor of Microsoft HoloLens as it provides improved graphics and improved hand movements tracking at a cheaper price. Figure 3 shows a user testing the Magic Leap-based station.

3.3 Software

The software application used in the experiments has been implemented using Unity3D 2019.1.0f2. This tool was selected because it makes easy deploying the same software for different devices. C# was chosen to program the scripts in Unity, using Visual Studio 2010 to code and debug them. 3DS Max and Adobe Photoshop were used to create the 3D model of the objects in the experiment. The software is run on the computers integrated in both HoloLens and Magic Leap systems, which, unlike other AR platforms come with dedicated computers to host and run all the software.

The AR application used in the experiment is composed of several modules (see Fig. 4). The *Manipulation Controller* is the component that manages the control related to the location and displacement of the objects in the scene. For example, when a user moves his/her hand to grab an object, an action is triggered and the object is moved by this component according to the corresponding movement of the hand. The *Tracking Controller* is responsible for identifying each action that the user performs during the session, so each time the user moves or drops an object, the action is recorded in a specific codification. The *AR Camera Controller*

Fig. 3 A person using the Magic Leap-based application at UALR

manages the cameras of the application. A camera in the application represents the user's point of view. Therefore, it is important to be able to control the rotation and translation of the camera depending on the device that is being used. Since we define two different instances for the two setups (HoloLens and Magic Leap), the AR Camera Controller is a dual component. The *Input Controller* is also a dual component. It has two different instances that handle the interaction of the two types of interaction systems used by the two different setups. The first
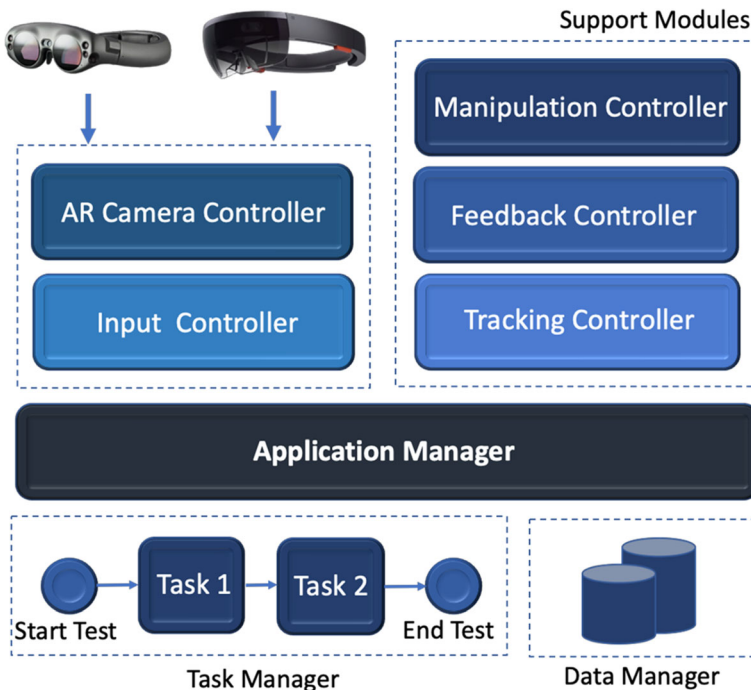

Fig. 4 Scheme of the AR application designed for the experiments

instance handles the data collected by HoloLens' sensors. The second instance handles the inputs to the Magic Leap device. Both instances perform a translation of these inputs to the events of the application. Gestures and motion are the input elements utilized in this setup. The *Feedback Controller* is responsible for the activation of visual and sound feedback according to the user's interaction in the application. For example, two different sounds are used to identify when an object has been grabbed or dropped respectively. The *Task Manager* is responsible for initiating and controlling the execution of the tasks defined in the application. The *Data Manager* is responsible for storing the data generated by each user while completing the assigned tasks. We record several objective datasets that represent the actions performed by the user. These datasets are saved in a CSV file for its registration and subsequent analysis. The list of datasets is detailed in the next section. Finally, the *Application Manager* is responsible for coordinating all other components. In other words, it is responsible for communicating the status of the application to the other components and for triggering the events or elements that must be raised to respond to the actions defined in the system.

4 Experimental study

As previously explained, the goal of the experiment is to analyze the use of natural hand interaction in AR systems by comparing the same application under two hardware setups that represent two different hand-tracking paradigms: a HoloLens-based system (representing a metaphoric interaction) and a Magic Leap-based system (representing an isomorphic interaction). The study tests if there are statistically significant differences in the use of hand interaction for two different tasks. Our hypothesis is that *there would be differences, in terms of both objective and subjective measures regarding the use of hand interaction between these two AR setups, in favor of the isomorphic setup*. Thus, several research questions are formulated and tested in this research in order to assess this initial hypothesis. These seven questions are enunciated next:

> Research Question 1 (RQ1): Which system will allow a more accurate completion of the tasks?
> Research Question 2 (RQ2): With what system will users complete the tasks with fewer mistakes?
> Research Question 3 (RQ3): Which system will allow a faster completion of the tasks?
> Research Question 4 (RQ4): Will the size of the virtual elements affect the performance of the users?
> Research Question 5 (RQ5): Which system will users consider more useful for these tasks?
> Research Question 6 (RQ6): Which system will users prefer?
> Research Question 7 (RQ7): Which system will users recommend?

4.1 Participants and procedure

In order to answer these questions and analyze differences between the two setups regarding hand interaction, we designed a set of task-based experiments. We ran our experimental study

in the Emerging Analytics Center (EAC) of the University of Arkansas at Little Rock (UALR). Since the goal of this research is to study the use of natural hand interaction in AR systems, we decided to recruit participants who did not have any previous experience using AR technologies with natural interaction, in order to avoid skilled users, which can benefit from some earlier experiences. The experiment was announced in every department of UALR. As a result, 45 people volunteered and registered to perform the experiment by signing up online. Of the 45 participants, 18 were women (40%) and 27 men (60%), with ages ranging from 18 to 61 (mean 34.67 ± 12.40).

In order to avoid biases for either of the systems, users were randomly assigned to two groups of roughly equal size: group A (24 participants) and group B (21 participants). Users in group A tested first the HoloLens-based system and then the Magic Leap, whereas users in group B tested first the Magic Leap-based system and then the HoloLens. The age, gender and group distribution of the participants are shown in Table 1.

The experiment was divided into two similar phases, one phase for each hardware setup. Therefore, users tested first one system, filled a questionnaire about that system, then tested the other system, filled a similar questionnaire about the use of the second system and finally filled a comparative questionnaire. Figure 5 summarizes the experimental protocol.

When a person arrived at EAC to perform the experiment, we carried out the following 8-step protocol, which includes the aforementioned two-phase design.

1 - Presentation and description. Before proceeding with the experiment, users were provided with a description about the tasks they had to complete and the maximum time available to complete the experiment (40 min in total, more than enough to complete all the required tasks). Then, users were required to sign a compulsory informed consent, where they declared to agree with the terms of the experiment, and fill a short questionnaire in order to provide some basic demographic information (gender, age and profession). They were informed that the application records performance data and that the experiment was completely anonymous.

2 - Instruction and practice. Before the start of the experiment, users received a short briefing on how to use the HoloLens or the Magic Leap (depending on which device they would have to test first). In both cases, a free practice of 5 min was carried out on three main actions: grab, drag and drop.

3 - Experiment. The experiment consisted in two different tasks, each one repeated three times (with three different conditions that will be explained in section 4.2), that the

Table 1 Age, gender and group distribution of the participants of the experiment

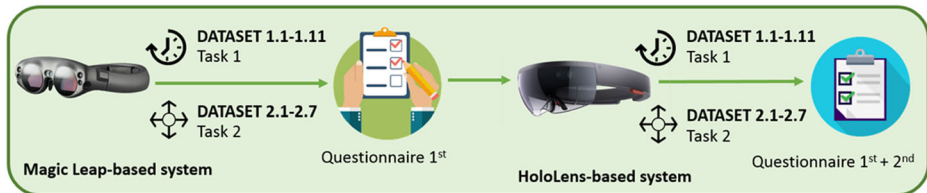| Age Group	<20	20–29	30–39	40–49	>50	Total
Males	3	9	10	3	2	27
Group 1	3	3	6	1	0	13
Group 2	0	6	4	2	2	14
Females	1	3	4	6	4	18
Group 1	0	2	2	4	3	11
Group 2	1	1	2	2	1	7
Total	4	12	14	9	6	45
Total (%)	8.9%	26.7%	31.1%	20.0%	13.3%	100%
Group 1	3	5	8	5	3	24
Group 2	1	7	6	4	3	21

Group A
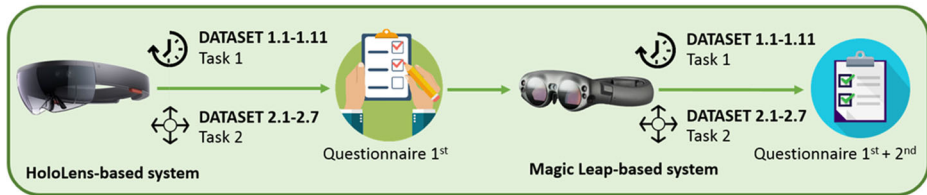
Group B

Fig. 5 Graphical description of the experimental protocol

participants needed to complete in the corresponding hardware setup. Each of the three trials of each of the two tasks had to be completed in no more than 120 s. Therefore, the maximum testing time was 12 min (two tasks, three times, 120 s each) per setup. During the experiment, user events were monitored in such a way that every meaningful action was measured and recorded. The tasks and the list of the datasets gathered by the application are explained in section 4.2.

4 - Setup evaluation. After users finished the tasks in the first setup, they were prompted to complete Questionnaire 1. Table 2 lists the questions asked in this subjective questionnaire. These were presented as 7-scale Likert questions with 1 meaning *strongly disagree*, 2 *disagree*, 3 *somewhat disagree*, 4 *neutral*, 5 *somewhat agree*, 6 *agree* and 7 *strongly agree*, except for the last three questions where 1 means *poor*, 2 *bad*, 3 *somewhat bad*, 4 *neutral*, 5 *positive*, 6 *good* and 7 *excellent*. Instead of analyzing the results of each question individually, the questions were grouped in six factors, as in [54]: sensory factors (SF), control factors (CF), distraction factors (DF), ergonomic factors (EF), realism factors (RF) and other factors (OF). These factors are adapted from the work described in [63]. There were also three additional questions about depth perception, usefulness and a global score. Therefore, nine datasets with subjective data, shown in Table 3, were created from the answers of this questionnaire.

Steps 5–7. Once a participant finished the experiment with the first hardware setup, all the previous steps of the process were repeated (except from the presentation step) using the other hardware setup.

8 - Final comparative evaluation. When the tasks in the second hardware station were also completed and users had completed Questionnaire 1 about the second setup, they were asked to fill Questionnaire 2, shown in Table 4, about user preference and recommendation regarding these two setups. The two-choice questionnaire included also an open-ended question "Additional comments and explanations" for the users to leave their impressions about the two setups or comments about the different answers provided for the three choice questions of this questionnaire.

Table 2 Questionnaire 1

Question (factor)

The information displayed on the device was adequate. (SF)
It was easy to handle the AR application. (CF)
The information displayed on the device was easy to read. (SF)
The information displayed on the AR device was clear. (SF)
It was easy to handle the device and its accessories. (CF)
I did not have to strive to recognize the instructional elements as 3D elements. (RF)
The 3D virtual elements looked like real. (RF)
The handling of the device and its accessories was simple and without complications. (CF)
The system responded to my actions adequately. (CF)
The handling of the system and its accessories was natural. (CF)
I did not feel delays between my actions with the device and the expected results. (CF)
The control mechanisms of the AR (glasses, lights, surface, etc.) did not distract me. (DF)
I got used to the AR in this device. (CF)
The device and the application were easy to use. (EF)
I found very useful the information provided by the AR application to complete the actions. (RF)
I had the impression that the aid elements appeared in 3D on the device. (RF)
The AR application helped me to complete the require actions. (RF)
I found the instructional elements to be useful. (RF)
I had the impression that the 3D labels, required for the task, were part of the scene. (RF)
There were moments that I thought that the elements that appeared on the device were real. (RF)
I did not pay attention to differences between the instructional elements and the actual device. (RF)
I had the impression that I could have touched the items that appear in the AR. (RF)
I liked the visual aids to help me complete the task. (RF)
I liked how virtual elements correlate with the actual device. (RF)
I have not felt any kind of discomfort during the experience (motion sickness). (EF)
I have felt the sensation of going in motion with this system. (OF)
I would like to use this technology with other uses. (OF)
The use of the system and its accessories was comfortable for my legs and arms. (EF)
I liked the experience of using a AR application. (OF)
The use of the system did not require a great effort from the legs or arms. (EF)
The use of the system did not require a great mental effort. (SF)
I have focused on the actions I had to do and not on the system or the environment. (DF)
My arms and legs are not tired after the experiment. (DF)
I felt involved during the experience. (OF)
At the end of the experience, I was an expert in the management of the system. (CF)
Rate the feeling of 3D (depth perception).
Rate the system as a device (global score).
Rate the usefulness of the system as a system to perform tasks that require precision skills (perceived uselfulness of the system).

Table 3 Datasets generated in the experiment from the answers to Questionnaire 1

Id	Description
SF	Sensory factors
CF	Control factors
DF	Distraction factors
EF	Ergonomics factors
RF	Realism factors
OF	Other factors
3D	3D depth perception
SC	Rated score of the system
US	Perceived usefulness of the system

Table 4 Questionnaire 2

#	Question
Q1-Q2	Which system did you find most useful? Why?
Q3-Q4	Which system did you like the most? Why?
Q5-Q6	What system would you recommend to be used as a part of related tasks? Why?

4.2 Tasks and objective datasets

As previously mentioned, two tasks were used to compare the two setups. Following the conclusions drawn in [11] and trying to avoid hand-tracking losses, we opted for one-handed tasks.

The first task is a pick-and-place task with three different levels of complexity. The main objective of this task is to evaluate how users can grab, move and place objects accurately using natural interaction (see Figs. 6 and 7). At the beginning of this task, each individual user was provided with eight virtual cubes placed on a real table. With these eight cubes, each user was expected to stack vertically at least five of the cubes to complete the task. The cubes need to be stacked on top of a blue-colored cube (see Figs. 6 and 7) that the system places automatically on the table at the beginning of the task. The cubes are physically simulated. Therefore, they collide (both with the real table and with other cubes) and fall if they are released too early or placed misaligned. For this reason, the participants had to be very careful when placing the cubes on the stack because they could fall down. Participants also had to avoid hitting the cubes that were already stacked when placing a new one on top of them. Participants were instructed to stack the cubes as perfectly aligned to the center of the stack - represented by the system-introduced blue cube – as possible. The level of complexity of this task is defined by the size of the cubes. For this reason, users need to complete the exercise three times with different cube sizes: small (S) cubes of size 7 cm, cubes of size 11 cm (M), and large cubes (L) of size 15 cm. Therefore, we perform three different experiments with this task.

The second task is a precision-movement drag-and-drop task with three levels of complexity, where users are required to grab a virtual sphere placed on a real table and move it through

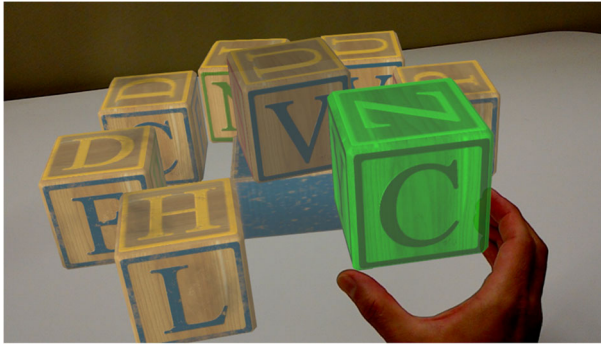**Fig. 6** Object picking (Task 1) from the user's point of view: metaphoric approach

Fig. 7 Object picking (Task 1) from the user's point of view: isomorphic approach

a static virtual pink curved path located just above the real table (see Fig. 8). Participants were instructed to move the ball through the path - starting from the right end and keeping it in contact with the path – and then drop it when the left end of the path was reached. If the ball goes off the path, it is detached from the participant's hand automatically by the application and falls down. Users can also drop the ball (intentionally or unintentionally) by releasing it. In all these undesired cases, participants can pick the ball from the table and try again to complete the task, although the elapsed time is not reset. The purpose of this task is to evaluate how accurately the sphere can be displaced – using natural interaction – along a curved path defined in 3D space, which allows also assessing the depth perception. The level of complexity of this task is defined by the size of the virtual sphere. For this reason, each user repeats this task three times with different ball sizes: S (5 cm in diameter), M (8 cm) and L (11 cm). The width of the curved path is fixed at 3 cm.

The advantage of these two tasks is that they are replicable, scalable, fast to test and can be analyzed with objective measures. They can be also considered generic tasks, yet sufficiently representative so that their results can be extrapolated to many specific AR-based contexts.

Since the HoloLens follows a metaphoric interaction paradigm, a pinch gesture is used to grab objects in the HoloLens-based setup. Thus, when the user's hand makes a pinch gesture close to a virtual object, this object is attached to the hand with a virtual fixed-joint. The object is released once the pinch gesture ends. It is important to point out that the pinch gesture is signaled by the HoloLens Application Programming Interface (API) and our software does not
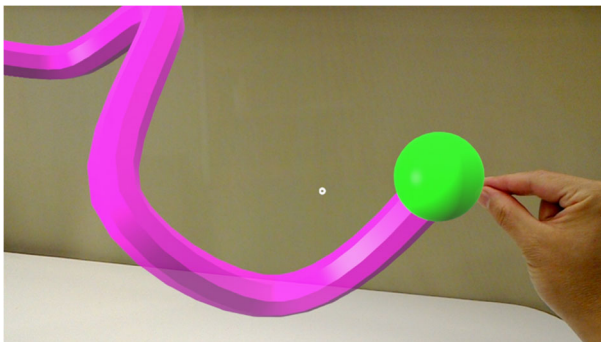

Fig. 8 Precision movement task (Task 2), from the user's point of view

have access whatsoever to the position of the fingertips in this setup. However, in the Magic Leap-based setup, the device API does provide access to the location of the fingertips. We use this information to grab objects using a more realistic isomorphic approach. In this setup, objects are grabbed by the detection of the collision of two of these fingertips (thumb and index fingers) with the geometry of the virtual objects. Once this collision occurs for both fingers, a virtual fixed-joint is created between these two fingertips and the object, which is considered grabbed until the fingertips move away from the virtual object's surface. The middle finger is also tracked in order to infer the orientation of the hand, but is not used for picking virtual objects. Using this technique, the physics simulation is more stable, since this kind of simulations get very unstable when a rigid virtual object is compressed by two or more virtual fingers (representing real fingers), since the absence of tactile or pressure feedback does not allow users to reduce the force when they feel that the object is firmly grasped. Therefore, we decided not to use a physics-based picking mechanism. Figure 6 (metaphoric, HoloLens) and Fig. 7 (isomorphic, Magic Leap) show these two different approaches from the user's point of view.

As it can be seen, the HoloLens-based system uses a simpler and less natural hand interaction system, but both systems are somehow semi-natural, since a real action with real objects would be performed in a similar, but nevertheless different, manner.

A large number of datasets are generated and stored automatically by the application while the users complete these tasks, so that it is possible to analyze their performance for each experiment. Table 5 lists the eleven datasets recorded automatically by the application for Task 1 and Table 6 shows the seven datasets recorded for Task 2. It is important to highlight that each dataset (of length 45) includes three sub-sets corresponding to the three sizes of each task.

5 Results and discussion

This section presents the results of the analysis performed using the data obtained in the different experiments. The first four research questions can be answered with the objective data gathered during the experiments. The rest of the research questions are analyzed with subjective data and with the two-choice comparative evaluation.

Table 5 Objective datasets generated in the experiment for Task 1

Id	Description	Best Value
1.1	Number of grabbed cubes	5–8
1.2	Number of dropped cubes	5–8
1.3	Number of stacked cubes	5–8
1.4	Number of unstacked cubes	0
1.5	Number of cubes stacked at the end	5–8
1.6	Total time to complete it (s)	low
1.7	Time holding any cube (s)	low
1.8	Average time each cube is stacked (s)	high
1.9	Average time each cube is held (s)	low
1.10	Average distance from stacked cubes to center (m)	0
1.11	Standard deviation of the distance from stacked cubes to center (m)	0

Table 6 Objective datasets generated in the experiment for Task 2

Id	Description	Best Value
2.1	Number of grabbed balls	1
2.2	Number of fallen balls (off path)	0
2.3	Number of dropped balls	0
2.4	Total time to complete it (s)	low
2.5	Time holding any ball (s)	low
2.6	Average distance from the balls to the path (m)	0
2.7	Standard deviation of the distance from the balls to the path (m)	0

5.1 Statistical analysis of objective data

In order to compare the two systems and test if there are objective significant differences between the two AR systems, we performed a statistical analysis with the data collected in the experiments, using IBM SPSS 26 software. All the statistical tests were two-tailed and were conducted at the 0.05 significance level.

First, we analyze the datasets measuring objective information about users' performance when completing the two tasks included the experiments. Table 7 shows the analysis of objective data (averaged for all the participants), for Task 1, for those participants who tested each of the systems (HoloLens or Magic Leap) first. Therefore, these participants are not influenced by a previous experience using the other system. Table 8 shows the same analysis for Task 2. No statistically significant differences are found for Task 1, whereas only one of the seven datasets (2.4, total time to complete the task) reveals a significant difference (in favor of Magic Leap) for Task 2.

In order to dig deeper into the question, we also compare the results by group. Table 9 (for Group A) and Table 10 (for Group B) show the results, for Task 1, of comparing the objective data within the two groups. Interestingly enough, four datasets show statistically significant differences in one direction (favoring Magic Leap) for Group A, whereas two datasets show differences for Group B but in the exact opposite direction.

Similarly, in the case of Task 2, shown in Table 11 (Group A) and Table 12 (Group B), there are four and three datasets, respectively, showing statistically significant differences, in favor of the Magic Leap-based setup and in favor of the HoloLens-based setup. Although these

Table 7 Study of statistically significant differences in objective data between the two setups, for Task 1 (averaged for the three sizes). Means, standard deviations (SD), unpaired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (B)	t	p	d
1.1	12.847±5.718	13.563±5.383	−0.431	.669	−0.129
1.2	12.931±6.364	13.468±5.402	−0.303	.763	−0.091
1.3	6.146±1.836	5.317±2.169	1.388	.172	0.414
1.4	3.292±1.827	2.476±1.884	1.472	.148	0.440
1.5	2.472±1.451	2.933±1.457	−1.048	.301	−0.317
1.6	94.973±27.722	88.890±24.588	0.774	.443	0.233
1.7	3.962±1.346	3.503±1.117	1.234	.224	0.373
1.8	12.124±13.416	14.987±16.157	−0.649	.520	−0.194
1.9	3.433±1.455	3.553±1.815	−0.247	.806	−0.074
1.10	0.022±0.006	0.022±0.008	0.163	.871	0.049
1.11	0.011±0.004	0.010±0.004	0.829	.411	0.247
Groups sizes= 24 (A), 21 (B). Degrees of freedom= 43					

Table 8 Study of statistically significant differences in objective data between the two setups, for Task 2 (averaged for the three sizes). Means, standard deviations, unpaired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (B)	t	p	d
2.1	4.271±5.451	2.143±1.769	1.710	.095	.589
2.2	5.278±10.028	1.397±2.105	1.738	.089	.640
2.3	6.396±10.027	2.563±2.396	1.707	.095	.617
2.4	51.186±37.876	25.912±21.156	2.656	.011	.856
2.5	12.629±7.550	15.325±8.033	−1.160	.252	−.346
2.6	0.028±0.005	0.027±0.003	0.878	.385	.275
2.7	0.006±0.001	0.006±0.001	0.070	.944	.021
Groups sizes= 24 (A), 21 (B). Degrees of freedom= 43					

Table 9 Study of statistically significant differences in objective data between the two setups, for group A and Task 1 (averaged for the three sizes). Means, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (A)	t	p	d
1.1	12.847±5.718	19.833±8.405	−3.421	.002	−0.350
1.2	12.931±6.364	19.833±8.405	−3.237	.004	−0.388
1.3	6.146±1.836	5.764±1.353	0.944	.355	−0.308
1.4	3.292±1.827	3.514±2.398	−0.422	.677	−0.537
1.5	2.472±1.451	5.125±1.766	−5.757	.000	−0.382
1.6	94.973±27.722	80.937±30.650	1.924	.067	−0.315
1.7	3.962±1.346	2.291±1.134	5.202	.000	−0.431
1.8	12.124±13.416	15.728±22.439	−0.588	.562	−0.963
1.9	3.433±1.455	2.291±1.134	−1.482	.152	−0.508
1.10	0.022±0.006	0.020±0.005	1.827	.081	−0.286
1.11	0.011±0.004	0.010±0.004	1.486	.151	−0.356
Degrees of freedom= 23					

paired comparisons are less relevant than those shown in Tables 7 and 8, since a certain learning effect can occur, the results show that the setup tested secondly gets better values. This is an indication that there are not important differences between the two setups.

Table 10 Study of statistically significant differences in objective data between the two setups, for group B and Task 1 (averaged for the three sizes). Means, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD Magic Leap (B)	Mean±SD HoloLens (B)	t	p	d
1.1	13.563±5.383	17.778±9.756	−1.601	.125	−0.557
1.2	13.468±5.402	17.810±9.791	−1.666	.111	−0.571
1.3	5.317±2.169	5.381±1.363	−0.118	.908	−0.036
1.4	2.476±1.884	3.127±2.649	−1.017	.321	−0.287
1.5	2.933±1.457	5.183±1.029	−5.269	.000	−1.810
1.6	88.890±24.588	77.449±30.273	1.580	.130	0.417
1.7	3.503±1.117	2.284±0.841	4.003	.001	1.245
1.8	14.987±16.157	10.472±11.494	1.283	.214	0.327
1.9	3.553±1.815	2.284±0.841	0.125	.902	0.956
1.10	0.022±0.008	0.020±0.005	1.019	.320	0.333
1.11	0.010±0.004	0.009±0.003	1.580	.130	0.509
Degrees of freedom= 20					

Table 11 Study of statistically significant differences in objective data between the two setups, for group A and Task 2 (averaged for the three sizes). Means, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (A)	t	p	d
2.1	4.271±5.451	5.417±6.288	−0.745	.464	−0.195
2.2	5.278±10.028	0.542±0.900	2.363	.027	0.867
2.3	6.396±10.027	5.958±6.910	0.195	.847	0.052
2.4	51.186±37.876	28.745±21.179	2.954	.008	0.718
2.5	12.629±7.550	7.170±4.304	4.527	.000	0.921
2.6	0.028±0.005	0.015±0.003	11.628	.000	3.241
2.7	0.006±0.001	0.007±0.001	−0.583	.565	−0.125
Degrees of freedom = 23					

With respect to research questions RQ1, RQ2 and RQ3 we need to analyze specific datasets for each task. Regarding RQ1, neither of the systems is more accurate than the other for neither of the tasks, since none of the datasets that measure accuracy (1.10, 1.11, 2.6, 2.7) show statistically significant differences in the unpaired comparison (Tables 7 and 8). A similar situation appears with respect to RQ2, since none of the datasets related with mistakes (1.4, 2.2, 2.3) show statistically significant differences between the two systems. The situation is different for RQ3, since from dataset 2.4 we can see that Task 2 can be completed faster with a Magic Leap. However, Task 1 is not completed faster in the Magic Leap-based setup.

Regarding RQ4, the previous statistical analyses are not appropriate to get an answer for this question. Since the question is about the effect of size in the performance of the users, a one-way analysis of variance (ANOVA) for the three sizes of virtual objects (S, M, L) is performed comparing the *total time* datasets (1.6 and 2.4) for those participants who tested each of the systems (HoloLens or Magic Leap) first. The results of the ANOVA test are presented in Table 13.

According to these results, no statistically significant difference in the time necessary to complete Task 2 can be attributed to the size of the balls. The results for Task 1 are different. For the HoloLens-based setup, no differences can be found. However, in the case of the Magic Leap, there are statistically significant differences (although the p value is just 0.049) for Task 1 that can be explained by the size of the virtual cubes. Therefore, the answer to RQ4 is that the size of the virtual elements does not play a central role in the performance of the users. There

Table 12 Study of statistically significant differences in objective data between the two setups, for group B and Task 2 (averaged for the three sizes). Means, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD Magic Leap (B)	Mean±SD HoloLens (B)	t	p	d
2.1	2.143±1.769	3.762±5.280	−1.506	.148	−0.459
2.2	1.397±2.105	0.159±0.343	3.032	.007	1.012
2.3	2.563±2.396	3.921±5.573	−1.254	.224	−0.341
2.4	25.912±21.156	32.722±17.892	−1.861	.078	−0.349
2.5	15.325±8.033	9.676±6.498	4.328	.000	0.777
2.6	0.027±0.003	0.013±0.002	17.387	.000	5.219
2.7	0.006±0.001	0.006±0.001	0.197	.846	0.056
Degrees of freedom = 20					

Table 13 One-way ANOVA for unpaired data. Sizes S, M, L are compared for the *total time* datasets (1.6 for Task 1, 2.4 for Task 2)

Task	Setup	Size	Mean±SD	F	p	η^2
1	HoloLens	L	93.363±36.794	0.469	.628	0.014
		M	92.512±34.465			
		S	101.381±32.055			
	Magic Leap	L	75.058±40.539	3.181	.049	0.098
		M	87.920±37.810			
		S	102.778±25.905			
2	HoloLens	L	27.977±13.537	0.652	.525	0.024
		M	34.323±24.712			
		S	29.747±10.156			
	Magic Leap	L	23.362±16.108	0.68	.935	0.002
		M	21.672±17.732			
		S	22.933±9.182			

Degrees of freedom= (2,71) - HoloLens, (2,62) - Magic Leap

are differences, but these differences are limited to only one of the tasks and just with one of the setups.

5.2 Statistical analysis of user responses

Next, we analyze the datasets of the responses to Questionnaire 1, which report the perceptions of the participants of the study, grouped by factors, in Table 14 (unpaired t-test comparing users who tested each system first), Table 15 (paired t-test for participants in group A) and Table 16 (paired t-test for group B).

As it can be seen, no statistically significant differences can be identified when comparing users testing each system first (Table 14). In addition, only two measures (RF, US) present statistically significant differences for the comparison within group A, and just one factor (OF) in the case of group B. Unlike the results obtained for the paired t-tests for objective data, in both cases (Tables 15 and 16) the differences in this analysis reveal that the increased value in the factors occurs in favor of the Magic Leap-based setup, which suggests a slight subjective slant for this setup. However, control factors (CF) and ergonomic factors (EF), which are related to interaction, do not show any significant differences.

Table 14 Study of statistically significant differences in user responses between the two setups. Means, standard deviations, unpaired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (B)	t	p	d
SF	6.042±0.934	6.024±1.277	0.054	.957	0.016
CF	5.104±1.327	4.970±1.386	0.331	.742	0.099
DF	5.333±1.255	5.667±1.101	−0.941	.352	−0.283
EF	5.417±1.371	5.619±1.420	−0.486	.629	−0.145
RF	5.441±0.959	5.476±1.225	−0.108	.914	−0.032
OF	5.792±1.010	5.917±1.035	−0.409	.684	−0.122
3D	5.917±1.139	6.000±1.483	−0.213	.832	−0.064
SC	5.375±1.555	5.667±1.461	−0.646	.522	−0.193
US	4.792±2.000	4.571±2.158	0.355	.724	0.106

Groups sizes= 24 (A), 21 (B). Degrees of freedom= 43

Table 15 Study of statistically significant differences in user responses between the two setups, for group A. *Means*, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD HoloLens (A)	Mean±SD Magic Leap (A)	t	p	d
SF	6.042±0.934	6.312±0.719	−1.320	.200	−0.327
CF	5.104±1.327	5.380±1.228	−0.847	.406	−0.216
DF	5.333±1.255	5.764±1.019	−1.501	.147	−0.379
EF	5.417±1.371	5.865±0.944	−1.853	.077	−0.387
RF	5.441±0.959	5.858±0.667	−2.507	.020	−0.512
OF	5.792±1.010	5.990±0.904	−1.024	.317	−0.207
3D	5.917±1.139	6.420±0.654	−1.958	.062	−0.558
SC	5.375±1.555	6.000±1.142	−1.760	.092	−0.460
US	4.792±2.000	5.830±1.274	−2.460	.022	−0.635
Degrees of freedom = 23					

5.3 Statistical analysis of two-choice questions

In this section, we analyze the responses obtained from Questionnaire 2. First, we analyze the two-choice questions (Q1, Q3, Q5) in which users are prompted to decide between the two setups regarding usefulness, preference and recommendation. As depicted in Table 17, the setup using Magic Leap is perceived more useful (60% vs 40%), preferred (62.22% vs 37.88%) and recommended (64.44% vs 35.56%) over the HoloLens-based setup. Within groups, the differences are similar, except in the case of Q5, in which a 70.8% of the participants in Group A recommends the setup they tested secondly (i.e., the Magic Leap-based setup).

A binomial test, however, reveals that the differences, despite being noticeable, are not statistically significant under this test, although in the case of Q5, by a very small margin. Therefore, research questions RQ5, RQ6 and RQ7 have not a clear answer. Nevertheless, it seems that the Magic Leap-based setup receives opinions that are more favorable than those obtained for the HoloLens-based system are. The analysis of the open-ended questions (Q2, Q4, Q6) of Questionnaire 2 adds some details to this matter. In question Q4, several users affirm that Magic Leap was "more natural" and "felt more in control". In this question, those who opted for the HoloLens commented that it was "easier to use". The response for Q2 does not offer much clarification since similar comments (such as "more responsive" or "easier to use") are found for justifying either choice. However, some users complained about the limited

Table 16 Study of statistically significant differences in user responses between the two setups, for group B. Means, standard deviations, paired t-test (t and p) and Cohen's test (d)

Id	Mean±SD Magic Leap (B)	Mean±SD HoloLens (B)	t	p	d
SF	6.024±1.277	5.714±1.233	1.593	.127	0.247
CF	4.970±1.386	4.827±1.587	0.411	.685	0.096
DF	5.667±1.101	5.238±1.403	1.536	.140	0.342
EF	5.619±1.420	5.060±1.398	1.762	.093	0.397
RF	5.476±1.225	5.270±1.452	1.098	.285	0.154
OF	5.917±1.035	5.214±1.379	3.038	.006	0.582
3D	6.000±1.483	5.380±2.037	1.652	.114	0.352
SC	5.667±1.461	4.950±1.717	1.826	.083	0.453
US	4.571±2.158	4.620±2.156	−0.094	.926	−0.023
Degrees of freedom = 20					

Table 17 Study of statistically significant differences in user responses for the two-choice questions

Question		HoloLens	Magic Leap	Binomial p value
Q1	Group A	10 (41.7%)	14 (58.3%)	
Usefulness	Group B	8 (38.1%)	13 (61.9%)	
	Total	18 (40.0%)	27 (60.0%)	.233
Q3	Group A	9 (37.5%)	15 (62.5%)	
Preference	Group B	8 (38.1%)	13 (61.9%)	
	Total	17 (37.8%)	28 (62.2%)	.135
Q5	Group A	7 (29.2%)	17 (70.8%)	
Recommendation	Group B	9 (42.9%)	12 (57.1%)	
	Total	16 (35.6%)	29 (64.4%)	.072

FOV of the HoloLens and about hand tracking loses with this setup. Finally, question Q6 also offers mixed comments and opposite choices are justified with similar comments (such as "more precision" or "easier to use"). Some users commented, however, that HoloLens provides clearer visuals, something that is not found in the comments of Magic Leap, despite having a wider FOV (something that some users also emphasize). The general conclusion that can be extracted from these comments is that there is a small preference for the Magic Leap but the differences are small and different subjects experience dissimilar, even contradictory, perceptions.

5.4 Statistical analysis of correlation and sources of variation

We have also analyzed the correlation between the factors extracted from the subjective responses (Questionnaire 1) given by the participants of each group. The importance of this type of test is that it allows measuring the consistence of the responses given by the participants. The results of this analysis include the significance levels and the correlation factors, which are shown in Figs. 9 and 10 (only those that are statistically significant) in colored circles with numbers. These numbers and colors represent the Pearson's correlation in 0–100 units. As it can be seen, the degree of correlation between the different factors and measures is really high, with correlation indexes above 0.6 between most measures. The correlation between the final score and most particular factors is also quite high. This means that the answers of the participants are consistent and reliable. Correlations are especially high in group B (with statistically significant correlations between each possible pair of factors, except for one).

Finally, we analyze if there is a significant interaction among the different features of the population and the data gathered for both tasks. In particular, we have considered the factors of gender, age, profession and tested system, for each dataset. A multifactorial ANOVA test revealed that there are not statistically significant differences between the two setups for these factors in Task 1. However, there are significant differences, in Task 2, between the five age groups considered (F [19, 37] = 20.769, p = .01, η^2 = 0.922). This is explained because older people tend to underperform with respect to younger people. The fact that it only appears in Task 2 is probably caused by the complexity of this task.

This age effect – and the small effect of the size of the virtual elements revealed by the one-factor ANOVA in section 5.1, Table 13 – are the only meaningful sources of variations in these experiments.
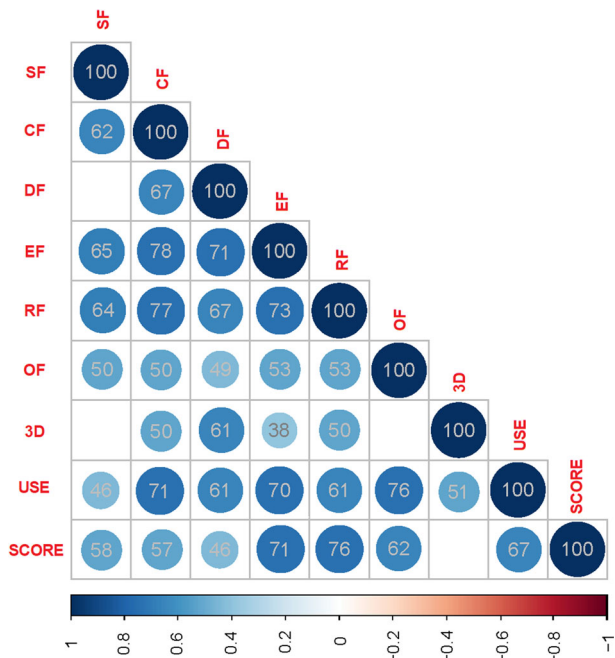
Fig. 9 Correlation plot for the questions in Questionnaire 1 - Group A
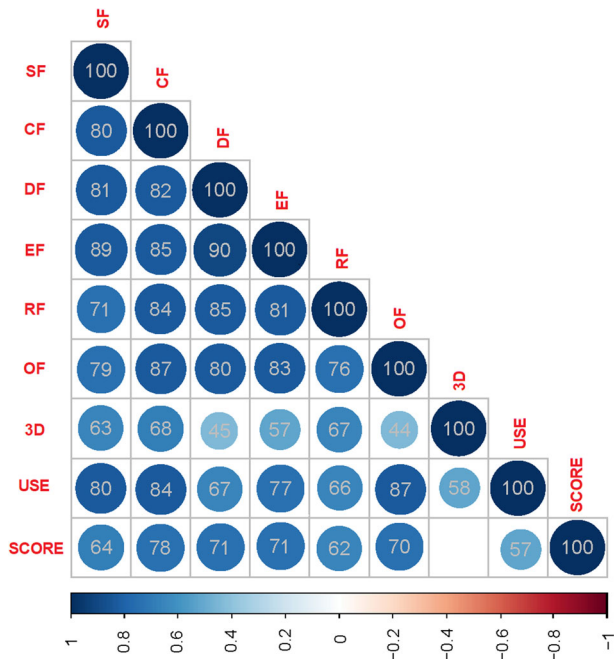

Fig. 10 Correlation plot for the questions in Questionnaire 1 - Group B

6 Conclusion and future work

Recent technological advances in AR, regarding mainly tracking and visualization with HMDs, have reshaped the field completely, closing the gap between research and the mass-market adoption of these devices. However, it is still unclear if more natural hand interaction systems provide benefits for AR applications. For this reason, we compare in this research article two optical-based AR systems: a HoloLens and a Magic Leap One. We do so by comparing natural hand interaction between the two setups. The former uses a metaphoric approach and represents the first of a series of optical devices capable of delivering AR experiences with head and hand tracking. The latter relies on an isomorphic approach, and represents the evolution of these type of systems since it provides more sophisticated hand tracking capabilities.

We raise several research questions regarding these two setups, which we answer after performing a series of task-based experiments using virtual elements that are moved using natural hand interaction. For this purpose, we collect objective data about user performance when completing these tasks. The results surprisingly show that there are very small differences in the use of hand interaction between these two systems, and the most recent one (represented by the Magic Leap One) is not significantly better in terms of accuracy and mistakes for the selected tasks. Only for one of the tasks do the users need more time to complete it using the HoloLens. There are also small differences with respect to the size of the virtual elements, which seems not to play a central role in the complexity of the task.

Besides the analysis of objective data, we also analyze several factors by means of a subjective questionnaire of user responses. No statistically significant differences can be found between users testing each of the systems for the first time, although a small edge in favor of Magic Leap can be identified in paired (within-groups) comparisons.

We also analyze the opinions of each of the users in terms of usefulness, preference and recommendation, by means of two-choice questions and open-ended questions where users can explain their choices. The results show that, although the Magic Leap-based system gets more support, the differences are not statistically significant under a binomial test.

The reason for this seemingly surprising result may lie in the uncanny valley in natural interaction systems explained in [38]. According to this idea, low-fidelity interaction systems that do not resemble real world interactions could provide better or similar user performance than more realistic interaction techniques. This might be the case in these experiments.

In order to reinforce this conclusion, further experiments with other setups should be performed. In addition, this work presents some limitations that should be taken in consideration. First, bimanual gestures are not considered in the experiments, even though they are allowed in both setups. Secondly, although the tasks have been carefully selected to be general and representative of real tasks in AR applications, the study is limited to two tasks and it is possible that deeper differences could be found using more specific tasks.

Nevertheless, the results obtained provide already useful information, since we honestly expected to find many more differences between the two setups (which are separated in time by more than a two-year difference) in this particular area, given the technological differences present between the two devices. In fact, our initial hypothesis, which is not confirmed by the results, was that there would be differences (in both subjective and objective measures), in favor of the isomorphic and newer setup (Magic Leap), in the use of hand interaction in the two AR tasks presented in the experiments.

Future work includes additional experiments with other optical-based AR solutions, such as HoloLens 2, or more sophisticated tasks, such as bimanual tasks, when devices with wider FOV allow doing so. It is also worth extending this research to other natural interfaces such as voice, eye tracking or body movement, and of course to the other two main problems in AR - visualization and tracking – in order to get a clearer picture of the maturity and practical usability of this technology.

Declarations

Conflict of interest   The authors declare no conflict of interest.

References

1. Affolter R, Eggert S, Sieberth T, Thali M, Ebert LC (2019) Applying augmented reality during a forensic autopsy—Microsoft HoloLens as a DICOM viewer. J Forensic Radiol Imaging 16:5–8
2. Al Janabi HF, Aydin A, Palaneer S et al (2020) Effectiveness of the HoloLens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study. Surg Endosc 34:1143–1149
3. Aliprantis J, Konstantakis M, Nikopoulou R et al (2019) Natural interaction in augmented reality context. In: VIPERC@ IRCDL, pp 50–61
4. Al-Kalbani M, Williams I, Frutos-Pascual M (2016) Analysis of medium wrap freehand virtual object grasping in exocentric mixed reality. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 84–93
5. Anderson R, Toledo J, ElAarag H (2019) Feasibility Study on the Utilization of Microsoft HoloLens to Increase Driving Conditions Awareness. In: 2019 SoutheastCon. IEEE, pp 1–8
6. Asgary A, Bonadonna C, Frischknecht C (2019) Simulation and visualization of volcanic phenomena using Microsoft Hololens: case of Vulcano Island (Italy). IEEE Trans Eng Manag 67:545–553
7. Azuma RT (1997) A survey of augmented reality. Presence Teleop Virt 6:355–385
8. Bai H, Sasikumar P, Yang J, Billinghurst M (2020) A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp 1–13
9. Bimber O, Raskar R (2005) Spatial augmented reality: merging real and virtual worlds. AK Peters/CRC Press

10. Borja EF, Lara DA, Quevedo WX, Andaluz VH (2018) Haptic stimulation glove for fine motor rehabilitation in virtual reality environments. In: International conference on augmented reality, Virtual Reality and Computer Graphics. Springer, pp 211–229
11. Chaconas N, Höllerer T (2018) An evaluation of bimanual gestures on the microsoft hololens. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 1–8
12. Chen Z, Li J, Hua Y et al (2017) Multimodal interaction in augmented reality. In: 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 206–209
13. Cruz E, Orts-Escolano S, Gomez-Donoso F, Rizo C, Rangel JC, Mora H, Cazorla M (2019) An augmented reality application for improving shopping experience in large retail stores. Virtual Reality 23:281–291
14. Cui N, Kharel P, Gruev V (2017) Augmented reality with Microsoft HoloLens holograms for near infrared fluorescence based image guided surgery. In: Molecular-guided surgery: molecules, devices, and applications III. International Society for Optics and Photonics, p 100490I
15. Dascano M (2018) Magic leap one AR: learning the essentials, pp. 27
16. Evans G, Miller J, Pena MI et al (2017) Evaluating the Microsoft HoloLens through an augmented reality assembly application. In: Degraded environments: sensing, processing, and display 2017. International Society for Optics and Photonics, p 101970V
17. Frutos-Pascual M, Creed C, Williams I (2019) Head mounted display interaction evaluation: manipulating virtual objects in augmented reality. In: IFIP Conference on Human-Computer Interaction. Springer, pp 287–308
18. Gibby JT, Swenson SA, Cvetko S, Rao R, Javan R (2019) Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. Int J Comput Assist Radiol Surg 14:525–535
19. Gimeno J, Casas S, Portalés C, Fernández M (2018) Addressing the occlusion problem in augmented reality environments with phantom hollow objects. Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) Munich, Germany, pp 21–24. https://doi.org/10.1109/ISMAR-Adjunct.2018.00024
20. Hanna MG, Ahmed I, Nine J, Prajapati S, Pantanowitz L (2018) Augmented reality technology using Microsoft HoloLens in anatomic pathology. Arch Pathol Lab Med 142:638–644
21. Hartholt A, Mozgai S, Fast E et al (2019) Virtual humans in augmented reality: a first step towards real-world embedded virtual roleplayers. In: Proceedings of the 7th international conference on human-agent interaction, pp 205–207
22. Hoover M (2018) An evaluation of the Microsoft HoloLens for a manufacturing-guided assembly task. Master Thesis, Iowa State University
23. Ishii H, Kobayashi M, Arita K (1994) Iterative design of seamless collaboration media. Commun ACM 37:83–97
24. Ishiyama H, Kurabayashi S (2016) Monochrome glove: A robust real-time hand gesture recognition method by using a fabric glove with design of structured markers. In: 2016 IEEE virtual reality (VR). IEEE, pp 187–188
25. Khalaf AS, Alharthi SA, Dolgov I, Toups ZO (2019) A comparative study of hand gesture recognition devices in the context of game design. In: Proceedings of the 2019 ACM international conference on interactive surfaces and spaces., pp 397–402
26. Kuhlemann I, Kleemann M, Jauer P, Schweikard A, Ernst F (2017) Towards X-ray free endovascular interventions–using HoloLens for on-line holographic visualisation. Healthc Technol Lett 4:184–187
27. Kyriakou P, Hermon S (2019) Can I touch this? Using natural interaction in a museum augmented reality system. Digit Appl Archaeol Cult Herit 12:e00088
28. Lee M, Billinghurst M, Baek W, Green R, Woo W (2013) A usability study of multimodal input in an augmented reality environment. Virtual Reality 17:293–305
29. Lee J, Jung S, Kim JW, Biocca F (2019) Applying spatial augmented reality to anti-smoking message: focusing on spatial presence, negative emotions, and threat appraisal. Int J Human–Comput Inter 35:751–760
30. Lia H, Paulin G, Yeo CT et al (2018) HoloLens in suturing training. In: Medical imaging 2018: image-guided procedures, robotic interventions, and modeling. International Society for Optics and Photonics, p 1057628
31. Ling FF, Elvezio C, Bullock J et al (2019) A Hybrid RTK GNSS and SLAM Outdoor Augmented Reality System. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 1044–1045
32. Liu H, Zhang G, Bao H (2016) Robust keyframe-based monocular SLAM for augmented reality. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 1–10

33. Liu Y, Dong H, Zhang L, El Saddik A (2018) Technical evaluation of HoloLens for multimedia: a first look. IEEE MultiMedia 25:8–18
34. Lu G, Shark L-K, Hall G, Zeshan U (2012) Immersive manipulation of virtual objects through glove-based hand gesture interaction. Virtual Reality 16:243–252
35. Macaranas A, Antle AN, Riecke BE (2015) What is intuitive interaction? Balancing users' performance and satisfaction with natural user interfaces. Interact Comput 27:357–370
36. Maniam P, Schnell P, Dan L, Portelli R, Erolin C, Mountain R, Wilkinson T (2020) Exploration of temporal bone anatomy using mixed reality (HoloLens): development of a mixed reality anatomy teaching resource prototype. J Vis Commun Med 43:17–26
37. McMahan RP, Alon AJD, Lazem S et al (2010) Evaluating natural interaction techniques in video games. In: 2010 IEEE symposium on 3D user interfaces (3DUI). IEEE, pp 11–14
38. McMahan RP, Lai C, Pal SK (2016) Interaction fidelity: the uncanny valley of virtual reality interactions. In: International conference on virtual, Augmented and Mixed Reality. Springer, pp 59–70
39. Microsoft (2016) Two hands gesture. Windows Mixed Reality Developer Forum
40. Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. IEEE Robot Autom Mag 19:98–100
41. Munsinger B, White G, Quarles J (2019) The usability of the Microsoft HoloLens for an augmented reality game to teach elementary school children. In: 2019 11th international conference on virtual worlds and games for serious applications (VS-games). IEEE, pp 1–4
42. NaturalPoint, Inc. (n.d.) Optitrack - Motion Capture Systems
43. Ni T, Bowman DA, North C, McMahan RP (2011) Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. Int J Hum-Comput Stud 69:551–562
44. O'Connor TF, Fach ME, Miller R, Root SE, Mercier PP, Lipomi DJ (2017) The language of glove: wireless gesture decoder with low-power and stretchable hybrid electronics. PLoS One 12:e0179766
45. Piumsomboon T, Altimira D, Kim H et al (2014) Grasp-Shell vs gesture-speech: a comparison of direct and indirect natural interaction techniques in augmented reality. In: 2014 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 73–82
46. Portalés C, Gimeno J, Casas S et al (2016) Interacting with augmented reality mirrors. In: Handbook of research on human-computer interfaces, developments, and applications. IGI Global, pp 216–244
47. Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, Amiras D (2018) Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2:2
48. Prilla M, Janßen M, Kunzendorff T (2019) How to interact with augmented reality head mounted devices in care work? A study comparing handheld touch (hands-on) and gesture (hands-free) interaction. AIS Trans Hum-Comput Interact 11:157–178
49. Quandt M, Hippert D, Beinke T, Freitag M (n.d.) User-centered evaluation of the learning effects in the use of a 3d gesture control for a mobile location based augmented reality solution for maintenance. Proceedings of DELbA 2020 Workshop on Designing and Facilitating Educational Location-based Applications co-located with the Fifteenth European Conference on Technology Enhanced Learning (EC-TEL 2020), Heidelberg, Germany
50. Rae E, Lasso A, Holden MS et al (2018) Neurosurgical burr hole placement using the Microsoft HoloLens. In: Medical imaging 2018: image-guided procedures, robotic interventions, and modeling. International Society for Optics and Photonics, p 105760T
51. Romanus T, Frish S, Maksymenko M, et al (2020) Mid-air haptic bio-holograms in mixed reality. arXiv preprint arXiv:200101441
52. Sadri S, Kohen S, Elvezio C et al (2019) Manipulating 3D anatomic models in augmented reality: comparing a hands-free approach and a manual approach. In: 2019 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 93–102
53. Seo DW, Lee JY (2013) Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. Expert Syst Appl 40:3784–3793
54. Serrano R, Morillo P, Casas S, Cruz-Neira C (2020) A comparative evaluation of a virtual reality table and a HoloLens-based augmented reality system for anatomy training. IEEE Transactions on Human-Machine Systems In Press. https://doi.org/10.1109/THMS.2020.2984746
55. Sørensen S, Jensen T (2019) Development of e-learning applications using HoloLens and mixed reality. In: European Conference on Games Based Learning. Academic Conferences International Limited, pp 649–656
56. Soto Ramos M (2019) Development of mixed reality applications using the magic leap one device. Departamento de Tecnología Electrónica y de las Comunicaciones, Universidad Autónoma de Madrid. http://hdl.handle.net/10486/688800

57. Stemasov E, Wagner T, Gugenheimer J, Rukzio E (2020) Mix&Match: towards omitting modelling through in-situ alteration and remixing of model repository artifacts in mixed reality. arXiv preprint arXiv: 200309169
58. Teng C-C, Redfearn B, Nuttall C et al (2019) Mixed reality patients monitoring application for critical care nurses. In: Proceedings of the third International Conference on Medical and Health Informatics, vol 2019, pp 49–53
59. Tian Y, Wang X, Yao H, Chen J, Wang Z, Yi L (2018) Occlusion handling using moving volume and ray casting techniques for augmented reality systems. Multimed Tools Appl 77:16561–16578
60. Valentini PP (2018) Natural interface for interactive virtual assembly in augmented reality using leap motion controller. International Journal on Interactive Design and Manufacturing (IJIDeM) 12:1157–1165
61. Vuletic T, Duffy A, Hay L, McTeague C, Campbell G, Grealy M (2019) Systematic literature review of hand gestures used in human computer interaction interfaces. Int J Hum-Comput Stud 129:74–94
62. Wang RY, Popović J (2009) Real-time hand-tracking with a color glove. ACM Transactions on Graphics (TOG) 28:1–8
63. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence 7:225–240
64. Xue H, Sharma P, Wild F (2019) User satisfaction in augmented reality-based training using Microsoft HoloLens. Computers 8:9
65. Yang C-K, Chen Y-H, Chuang T-J et al (2019) An augmented reality-based training system with a natural user interface for manual milling operations. Virtual Reality 1–13
66. Zhang W, Zhang Q, Liu Z et al (2020) Assembly method for satellite device installation based on HoloLens. In: IOP conference series: materials science and engineering. IOP Publishing, p 012020
67. Zuo Y, Jiang T, Dou J, Yu D, Ndaro ZN, du Y, Li Q, Wang S, Huang G (2020) A novel evaluation model for a mixed-reality surgical navigation system: where Microsoft HoloLens meets the operating room. Surg Innov 1553350619893236:193–202