

Document downloaded from:

<http://hdl.handle.net/10251/166906>

This paper must be cited as:

Morillo, P.; García García, I.; Orduña, JM.; Fernández, M.; Juan, M. (2020). Comparative study of AR versus video tutorials for minor maintenance operations. *Multimedia Tools and Applications*. 79(11-12):7073-7100. <https://doi.org/10.1007/s11042-019-08437-9>



The final publication is available at

<https://doi.org/10.1007/s11042-019-08437-9>

Copyright Springer-Verlag

Additional Information

Comparative study of AR versus video tutorials for minor maintenance operations

Pedro Morillo¹ · Inmaculada García-García² · Juan M. Orduña¹ · Marcos Fernández¹ · M. Carmen Juan²

Abstract

Augmented Reality (AR) has become a mainstream technology in the development of solutions for repair and maintenance operations. Although most of the AR solutions are still limited to specific contexts in industry, some consumer electronics companies have started to offer pre-packaged AR solutions as alternative to video-based tutorials (VT) for minor maintenance operations. In this paper, we present a comparative study of the acquired knowledge and user perception achieved with AR and VT solutions in some maintenance tasks of IT equipment. The results indicate that both systems help users to acquire knowledge in various aspects of equipment maintenance. Although no statistically significant differences were found between AR and VT solutions, users scored higher on the AR version in all cases. Moreover, the users explicitly preferred the AR version when evaluating three different usability and satisfaction criteria. For the AR version, a strong and significant correlation was found between the satisfaction and the achieved knowledge. Since the AR solution achieved similar learning results with higher usability scores than the video-based tutorials, these results suggest that AR solutions are the most effective approach to substitute the typical paper-based instructions in consumer electronics.

Keywords Augmented Reality · Comparative study · Real user study · Multimedia-based learning · Equipment maintenance

This work has been supported by Spanish MINECO and EU ERDF programs under grant RTI2018-098156-B-C55.

C> Juan M. Orduña
Juan.Orduna@uv.es

Inmaculada García-García
ingarcia@dsic.upv.es

¹ Departamento de Informática, Universidad de Valencia, Valencia, Spain

² DSIC, Universitat Politècnica de València, Valencia, Spain

1 Introduction

The need for green solutions that contribute to sustainable industry procedures has led consumer electronics industry to avoid paper-based user manuals, offering multimedia solutions like video-based tutorials as alternative guides to traditional manuals for minor repair and maintenance operations. These video tutorials can be made by the manufacturer, the reseller, or even the users/technicians themselves.

On other hand, Augmented Reality (AR) [4] has become a mainstream technology in the development of solutions for different fields [1, 21, 22], and also for repair and maintenance operations [5, 17, 37, 48, 51, 53]. However, one of the main difficulties for a definitive establishment of Augmented Reality systems in industrial environments is that their performance evaluation should be carried out in a trustworthy way, together with its competitors and in the same working environment where the AR system is going to be used. The literature review of the last decade on training on maintenance operations in industrial environments shows that the contributions typically compare the effectiveness of the AR-based training systems to paper-based traditional systems [14, 51]. Moreover, beyond their improvement over paper-based system, one of these studies shows that the use of AR-based systems achieves a reduction up to 75% in the error rate made by the workers using these systems [48], improving even the error rate achieved by Virtual Reality [17].

In the particular field of consumer electronics, Video-based Tutorials (VT) offer a very intuitive mechanism to show the working of the device to the final user, which can even perform some minor maintenance tasks on the device. In this sense, dynamic video content has been shown to offer a significant advantage over static media, like paper-based manuals [19]. The making of video tutorials has become a simple process even for domestic users, which are the main consumers of this kind of multimedia content. Some studies show that more than one third of the population older than 14 usually use video-based tutorials found on video platforms like Youtube or Vimeo for solving problems of daily life [40]. Given the daily use of consumer electronics, the effectiveness of the available guides for minor repair and maintenance operations of these devices can have important effects on both their average life expectancy and the customer satisfaction. Additionally, the worldwide extension of many consumer electronics devices highlights the need for effective multimedia guides. Nevertheless, to our knowledge, no studies comparing the benefits of AR versus video-based tutorials have been published. In particular, we focus on the typical challenge that arises in big (public or private) organizations with a large computer infrastructure (typically based on the same computer platform) on which users can perform minor maintenance operations. The organization must train its employees in these operations, because employees usually get confused with the short documentation provided by the manufacturer, which does not include repairing/maintenance instructions.

In this context, we carried out a study where, using a real situation, we compare the effectiveness of an AR mobile application to the effectiveness of video-based tutorials for performing minor maintenance operations on a well-known consumer electronic device (a small size computer). Also, we evaluated the usability of the system and the user satisfaction with this application, using a population of forty final users. In this study, our primary hypothesis was that the effectiveness of using our AR application would be at least as good as using video-based tutorials. Since the user satisfaction with “Augmented Reality” and “Video Tutorials” technologies favor the learning process in front of traditional technologies like “paper-based manuals” [8, 12, 27–29, 55], our secondary hypothesis was that the user satisfaction with the AR application would be higher than with the video tutorial. In turn, it suggested a third hypothesis: in case of having to select a single option to be implemented

in the maintenance program for the IT equipment of a large organization, the participants in the experiment would select the mobile AR application among other alternatives.

The rest of the paper is structured as follows. Section 2 shows the related work in the evaluation of AR systems for minor maintenance operations. Section 3 describes in detail the AR application developed for training users in two different maintenance tasks on a mini tower PC. Next, Section 4 explains all the aspects of the study carried out with real users. Section 5 presents and analyzes the comparative results of the experiment. Finally, Section 6 presents the main concluding remarks of the comparative study.

2 Related work

Since the development of Augmented Reality, it has been used in many industrial procedures for maintenance and repair tasks, and the effectiveness of this kind of tools has also been studied. The first assessment of the effectiveness of AR systems for assembly purposes (in industrial environments) [48] compares the use of three instructional media (paper-based manual, instructions using a monitor and instructions using a head-mounted display) with an AR system. The results indicate that overlaying 3D instructions on the actual work pieces reduced the error rate for an assembly task by 82%, particularly diminishing cumulative errors.

Yuan et al. [57] described the assembly domain as one of the most promising applications of AR. In this sense, authors claim that alternating the attention between the object to maintain and the instructions would consume valuable time. Although these concepts could be valid also for other fields of applications, Yuan focused his research on the development of a virtual interactive tool for supporting AR, and not on the user experience, as done in [39].

Radkowski et al. [41] analyzes different types of visual features for different assembly operations using AR systems. In order to gain an advantage from AR, the visual features used to explain a particular assembly operation must correspond to its relative difficulty level. The final goal is to associate different types of visual features to different levels of task complexity.

Gavish et al. [17] perform an evaluation study where forty expert technicians were randomly assigned to four training groups in an electronic actuator assembly task: VR (training with the VR platform twice), Control-VR (watching a filmed demonstration twice), AR (training with the AR platform once), and Control-AR (training with the real actuator and the aid of a filmed demonstration once). A post-training test evaluates performance in the real task. Results demonstrate that, in general, the VR and AR training groups required shorter training time compared to the Control-VR and Control-AR groups, respectively. These results suggest that the use of the AR platform for training industrial maintenance and assembly tasks should be encouraged.

Other authors [52] have developed a modular software framework for intelligent AR training systems, and a prototype based on this framework that teaches novice users how to assemble a computer motherboard. However, the experimental result is weak, since it includes a low number of participant (sixteen), the learning factor (determining if using one of the two compared systems has some effects on the scores for the second) is not included, and a statistical analysis (based on ANOVA or similar statistics) is not performed in the comparative study.

Sanna et al. [45] have developed an AR system based on handheld devices oriented to maintenance tasks on consumer devices. The authors decided to show the description of the task in the bottom of the display and provide a few buttons to navigate through the

procedure. Virtual animations are overlaid on the real environment at each step. This step-by-step approach [4] is the same that was presented at [18] using an authoring tool. However, the experimental result is also weak: a low number of participants are considered, the learning factor is not included, and a statistical analysis is not performed in the comparative study. Due to the nature of the participants, they claim that the obtained results suggest that AR benefits are dependent on participant's skills.

A comprehensive review of the AR-based assembly systems developed since 1990 until 2015 is presented in [51]. These systems are divided in the following categories: AR assembly guidance (the most frequent), AR assembly training (the most emerging), and AR assembly simulation systems. Authors emphasize that the bottlenecks for current AR assembly systems are, among other reasons, low intuitive user system interfaces, calibration requirements and uncomfortable user devices.

Other work [14] analyzes more than 20 reports about the effectiveness of AR for education, training, and performance purposes. This analysis states that, although AR applications may substantially improve human performance, more emphasis on empirical assessment of these applications is needed and recommended.

Uva et al [50] have recently proposed a SAR (Spatial Augmented Reality) system for maintenance and assembly operations. Authors present a prototype, where the technical information on a motorbike engine during a seven-task maintenance procedure was projected to the user. Moreover, they complete an experimental study (including learning factor) with sixteen participants to measure the user task performance (completion times and error rates) and to collect subjective evaluation. However, around 40 participants are typically recommended to achieve statistical robustness in this type of studies [20, 32, 34].

Another recent surveys [23, 39] with very similar conclusions have analyzed more than 30 main contributions in the field of AR for industrial purposes published between 1997 and 2017. They claim that AR technologies are immature for complying with industrial requirements of robustness and reliability, and there are no common AR architectures or standards to be applied in maintenance tasks.

Finally, some studies have combined videos and AR applications creating new multimedia content, which is a mixture of both alternatives. A remarkable system transfers automatically printed technical documentation, such as handbooks, to three-dimensional Augmented Reality applications [24]. Although the new "augmented" documentation works with minimal user input, the system requires the CAD model or 3D scan of the object described in the documentation, constraining the use to controlled environments. On the other hand, the authors do not compare their approach to the current video-tutorials or common AR applications, in order to show the performance of the proposed solution. Another approach uses smartphones to record and replay video content composited in-situ with a live view of the real environments [30]. This new content can be inserted in mobile AR applications using a panorama-based tracking approach. Unlike our sphere of working, this new multimedia content could be appropriate for outdoor tutorials where the user's body movements could need to be emphasized. Finally, a recent work presents a technique for retargeting conventional, two-dimensional videos into three-dimensional AR tutorials [31]. This approach merges video-based tutorials and AR applications and can be applied to many styles of video tutorials. The authors focus on tutorials oriented to altering the surfaces of the objects, such as soldering, make-up of decoration. The highly elaborated retargeted 3D tutorials need much work to be completed and are typically not oriented for maintenance, assembly or repair tasks in industrial environments.

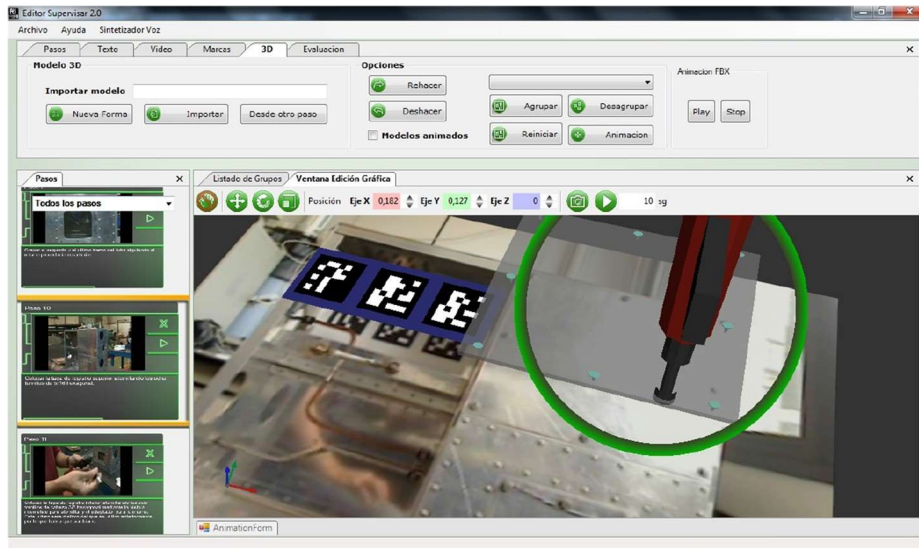


Fig. 1 Example snapshot of the SUGAR user interface

3 Materials and methods

In this section, we describe the AR application developed for helping in the learning process, including the software and hardware resources involved.

3.1 Description of the AR application

We developed an ad-hoc AR application for helping in the process of substituting some components of a PC. This development was based on SUGAR [18], an open-source software platform designed to enable a rapid prototyping of low-cost AR systems based on steps. SUGAR stands for System for the development of Unexpensive and Graphical Augmented Reality application. It is oriented to develop complex AR software applications based on procedural simulations, which are modeled following an easy-to-use AR authoring editor. The main purpose of this framework software is to significantly reduce the time required to develop Augmented Reality applications, regardless of the tracking technology used, i.e., marker-based, markerless-based, or even based on locating AR features [15]. This AR editor generates an exchange file, describing the AR procedure, which can be loaded into different AR devices not requiring high computational power. Using this tool, we developed an AR application consisting of the replacement of some components of a PC. This AR application allows for the performance of two different minor maintenance tasks on a mini tower PC. Figure 1 shows an example snapshot of the SUGAR interface when developing an AR application for industrial maintenance tasks.

Figure 2 shows the exploded view drawing in the assembly guide of the well-known ARTIGO A1000 computer, which is based on a ITX-Peak architecture and is commercialized as a mini tower PC. These images, obtained from the paper-based instructions delivered with the equipment, show the assembly of this computer. These instructions show that the substitution and maintenance tasks of components in this equipment could be simple, thanks

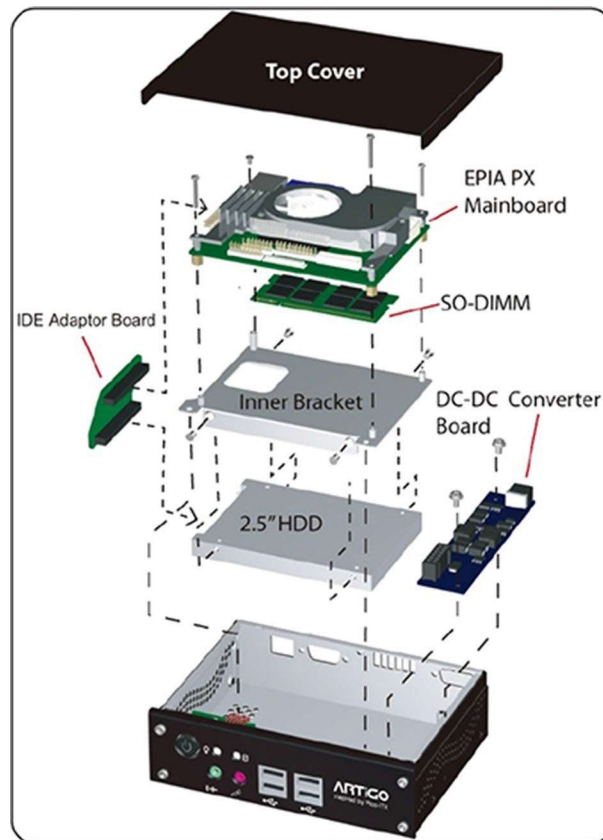


Fig. 2 Exploded view drawing in the assembly guide of the ARTIGO A1000 computer

to the modular configuration and layered structure configuration of the system. However, this image does not show that most of the elements are interconnected through multiple wires and connectors. This added complexity, together with the small size of the equipment (resulting in a very reduced workspace of 5.9×1.8 inches), makes any maintenance task uncomfortable and tedious. Due to this reason, we have defined the first task (labeled as 'simple procedure' in the smartphone apps and denoted as 'Assembly 1' in Section 5) to consist of the replacement of the RAM memory module in the mini tower PC.

Using the exploded view drawing in the assembly guide, we have sequenced this replacement in thirty steps, and we have modeled them using the SUGAR framework. As an illustrative example, Fig. 3 shows the materials and arrangement required for the edition and assembly of one of these steps, corresponding to the extraction of the faulty RAM module. This step requires four 3D models, corresponding to the motherboard and the RAM module, as well as two types of screwdrivers needed for the extraction operation. Of special note is the importance of the animations in many of the steps of the maintenance procedures. For example, in this case it is necessary to visually instruct the user to completely rotate the motherboard in order to find the RAM module, or to extract the RAM module through a gentle push-up of its grab tabs. SUGAR includes a simple interface which allows the implementation of simple affine transformations (rotations, translations or grading) for the

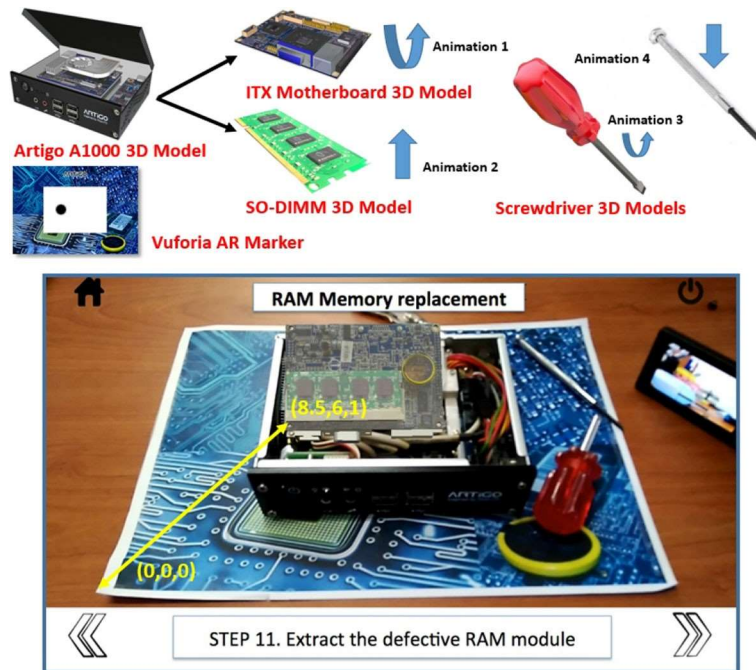


Fig. 3 Composition and required materials for editing and mounting step 11 (extracting the faulty RAM module)

3D models loaded. The development of each of the steps included in the AR-based tool for maintenance procedure did not take more than thirty minutes, since all the required 3D models were available in ‘3D Warehouse’, a well-known, open-source, 3D-model repository. Figure 3 also shows how the registration procedure is carried out (setting the relationship between the distances in the real world and the virtual objects). In each step, the user must indicate the distance (in cm.) between the bottom-left corner of the AR marker and the point where it will be shown the zero coordinates of the object. In this case, the zero coordinates of the object correspond to the bottom-left vertex of the 3D model simulating the RAM module. Since this point is a distance of (8.5,6.0,1.0) (along the (X,Y,Z) axis) to the bottom-left corner of the AR marker, this setting will guarantee a perfect merging between real image and computer-generated 3D models.

The complete maintenance procedure requires the removal of the external cover, disconnection of the DC-DC converter board (the system power supply), screwing out the retaining screws of the RAM module board to be replaced, and carefully extracting the module with a vertical movement. Next, after inserting the new RAM module, the same steps should be repeated in the reverse order, until completely closing the external cover and checking the correct working of the computer. The second task (labeled as ‘complex procedure’ in the smartphone apps and denoted as ‘Assembly 2’ in Section 5) consists of the substitution of the power supply module in the same equipment. Since this component is located at the innermost part of the computer, as shown in Fig. 2, the user should perform most of the steps in the simple procedure, remove most of the equipment connectors, remove the VGA output from the built-in Front Panel Board, and extract the EPIA PX Mainboard.

In order to make a fair comparison between the two maintenance support systems, we have also developed two complete video tutorials. These tutorial visually show and explain (besides labeling the text on the images) each of the respectively ten and thirty steps of the two maintenance procedures. The content, structure and actions to be carried out in each step have been taken from the maintenance manual issued by 'VIA Technologies, Inc.' through its web page.

3.2 Hardware and software

We have used Dolphin Player [13], a popular open-source audio and video player for Android devices, to customize our mobile video browser app. In addition to the basic features provided by this open source player (such as play, pause, stop, fast-forward and rewind), we have added two features. The first feature is, the generation of an acoustic alarm when the user makes the triple-tap gesture. The second feature is the generation of two folders with the videos, one of them secured by a password. We made the adaptation of the audio and video player using Android Studio version 2.3 as a cross-platform integrated development environment. We used SUGAR [18] to develop and run the Android mobile application including AR capabilities with extended tracking [7]. Extended tracking allows the development of continuous visual recognition even when the target/marker leaves the field of view of the camera or it becomes occluded. In order to achieve these features, we updated SUGAR to be compatible with Unity 3D and Vuforia. The new SUGAR update includes the development of scenes on ARCore (oriented to high-end Android devices), ARKit (oriented to iOS devices) or Vuforia Engine's VIO, all of them natively integrated in this version of the Vuforia SDK software platform [6]. The utilization of each of the three alternatives will depend on the hardware features of the mobile devices where the Augmented Reality application is going to be executed, whose installing and configuration procedure is performed transparently to the user. We have not used extended tracking in the experiments because, as Fig. 3 shows, we opted for a marker integrated in the scene, in order to focus the participants on the working area during the assembly exercises. While the participants are using the AR mobile application, this marker remains always within the field of view.

As an illustrative example, Fig. 4 shows a snapshot of a given moment in the experiment, showing a user at the moment when he is using the mobile video browser app to watch video number twelve of the 'complex procedure'. The smartphone is close to the mini tower PC on which the substitution tasks should be done. Figure 4 also shows the external case that was used to protect the smartphone from falls and shocks.

Figure 5 is a snapshot of another moment of the experiment, when the user is using the AR app. The figure shows the AR displayed in the smartphone display when it is focused on the mini tower PC.

4 Description of the study

This section describes all the elements in the study carried out: the participants involved, the measurements that were collected in the study, and the procedure that was followed to carry out the study. For each participant in the study, there was a staff person in charge of controlling the test. This staff person helped the participant in the test procedure and observed the making of the tests, measuring different metrics (observation measurements). The participants should carry out the two tasks related to PC computers maintenance described in Section 3.1. These assemblies were based on the configuration of the motherboard, also



Fig. 4 Smartphone displaying video number twelve of the complex procedure

denoted as shape factor, of type ITX-Peak. Concretely, we used ARTIGO A1000 (CPU VIA C7 1GHz, 1GB RAM DDR2, HD 160GB EIDE, VGA 1600x1200) platforms, which are widely used at industrial levels in office IT environments due to their excellent trade-off among cost, performance and size.

4.1 Participants

We have carried out a study involving forty people, in order to obtain statistically significant results [20]. From these forty people, twelve of them were women (30%) and twenty-eight

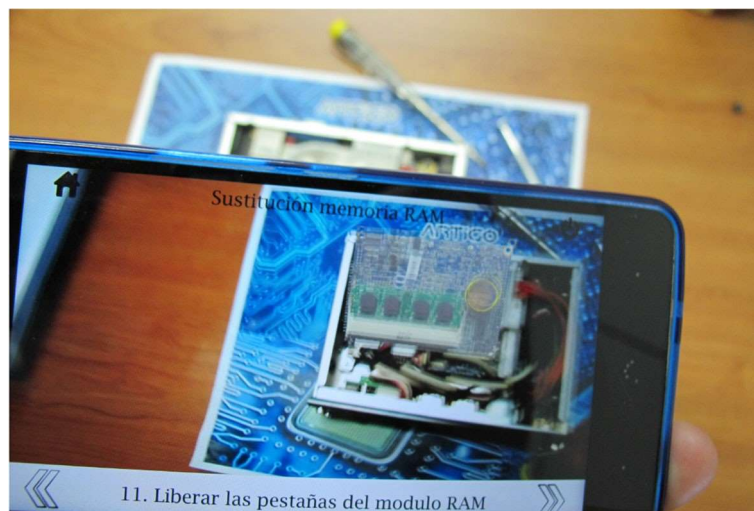


Fig. 5 Smartphone displaying AR images when focused on the mini tower PC

men (70%). The participants age ranged between twenty and fifty-four. The average age and standard deviation was 27.15 8.49. We split the participants into two groups of twenty people (denoted as groups A and B), randomly assigning the participants to each group. Each group was composed of six women and twelve men. From the forty people participating in the study, ten people (25%) had not any a priori computer skills (they had not any degree nor occupation related to computers). The remaining thirty people (75%) were either professionals working on any computer-related fields (fourteen people, 35%) or they were studying Computer Engineering or a similar Degree (sixteen people, 40%).

4.2 Measurements

Different metrics were measured during and after the making of the assemblies. The measurements came from the participants and from the staff (observers) through questionnaires that they should fill out at different moments, as described below. In particular, we used 6 questionnaires:

- PreTest: This questionnaire allows to determine the participant knowledge before using any learning method. It contains questions about the hardware present in computer systems. It also asks the user to determine, within a Likert scale ranging from 1 to 7 [43], his/her computer skills (Q18), and more concretely his/her PC repairing skills (Q19).
- Assembly 1 Observation: This questionnaire was completed by staff while the participant carries out the test. It measures, in a practical way, if the participant has understood the concepts related to the replacement of PC components in small size platforms of type ITX-Peak. While the participant is performing the test, the staff person annotates the degree of objective completion, as well as the quality of the assembly (OE1) using a Likert Scale. The observer also measures the time required to perform the test (OT1).
- PostTest1: The participants completed this form after using the first method. The questionnaire contains the same questions appearing in the PreTest questionnaire.
- Assembly 2 Observation: This questionnaire was completed by the staff person while the participant carries out the second test (see Table 5), annotating the corresponding measurements (OE2 and OT2).
- PostTest2: The participants completed this form (see Table 4) after using the second method. The questionnaire contains the same questions as in the PreTest questionnaire.
- Usability and Satisfaction: The participants completed this questionnaire after performing the two tests, therefore having used both learning methods. In this questionnaire, the participant expressed his/her opinion about different aspects of the application, like 3D realism, comfort factors, etc..

Since the three questionnaires PreTest, PostTest1, and PostTest2 measure the opinion of the participants about their knowledge on PC maintenance, they contain the same questions. The only difference among them is that the participants completed each of them at different moments of the procedure. Table 1 shows these questionnaires related to the participant's self-perceived knowledge, while Table 2 shows the usability questionnaire. It must be noted that additionally the experiment included observing staff that objectively evaluated the quality of the assemblies.

It must be noted that these questionnaires are basically positive, like many others used in problems on similar topics where validity or reliability of instruments are tested [33, 35, 38, 46, 56]. One of the fundamental reasons for the use of questionnaires with positive evaluations is that they are simple to understand by the participants, while the literature has

Table 1 PreTest, PostTest1, and PostTest2 questionnaires

Q1	I feel comfortable using video-tutorials for mounting or repairing some devices.
Q2	I could perform the typical repairs of a PC computer without the help of manuals.
Q3	I know the most important parts a PC CPU.
Q4	I can distinguish the basic elements of a PC with a Mini-ITX or Pico-ITX architecture.
Q5	I can explain the utility of the RAM memory in a PC.
Q6	I can identify the location of the RAM modules within a PC once the computer case has been opened.
Q7	I am able to replace the PC RAM memory modules by other similar modules.
Q8	I know the utility of a power supply/board for a PC.
Q9	I can identify the power supply/board inside a PC.
Q10	I am able to replace a computer PC power supply/board by other similar supplies/boards.
Q11	I could perform basic computer maintenance tasks with the help of a tutorial.
Q12	I can distinguish the most important elements of a PC once the computer case has been removed.
Q13	I consider the replacement of the most important parts of a PC as a simple task.
Q14	I can distinguish the location of the RAM modules from the secondary storage (hard disk drives) of a PC.
Q15	I am aware that all the connectors inside a PC are equal or very similar.
Q16	All the replacement tasks of PC components have the same complexity.
Q17	The replacement of a PC RAM module is simpler than the replacement of its power supply.
Q18	Mark your knowledge on informatics.
Q19	Mark your knowledge on PC maintenance.

shown that it does not compensate to correct the acquiescence bias. Indeed, some publications prove with multitrait-multimethod matrix (MTMM) that negative wording generates method effects [10, 49]. That is, the correction has more effect than using items in totally positive or totally negative questionnaires.

4.3 Procedure

Each of the participants in the experiment started his/her trial first reading written instructions (in paper) where it was explained how to interpret the information provided by the evaluated learning systems (either the mobile AR App or the mobile video browser App, installed in the smartphone used for the experiments). Next, a demonstrative video showed

Table 2 Usability questionnaire

Q1	The use of the application did not require a great mental effort.
Q2	The information displayed in the screen was adequate.
Q3	The information displayed in the screen was easy to read.
Q4	The information displayed in the screen was clear.
Q5	The use of the smartphone did not require a great effort of the arms.
Q6	The use of the smartphone was comfortable for my hands and arms.
Q7	The handling of the smartphone was easy.
Q8	I haven't felt any dizziness during the experiment.
Q9	My hands and arms did not get tired.
Q10	It was easy to control the application.
Q11	At no time did it seem to me that the smartphone was going to fall out.
Q12	The handling of the application was uncomplicated and simple.
Q13	The application reacted properly to my actions.
Q14	The handling of the application was natural.
Q15	I did not notice delays between my actions and the expected results.
Q16	The control mechanisms did not distract me.
Q17	I got used to the application at once.
Q18	The application was easy to use.
Q19	It seems to me very useful the information I was provided with.
Q20	I had the feeling that the help elements appeared on the device.
Q21	The application helped me to find the required elements of the device.
Q22	I found very useful the help elements.
Q23	I had the feeling that the help elements were part of the scene.
Q24	The elements displayed on the device have helped me in my task.
Q25	There were some moments when I thought that the elements appearing on the device were real.
Q26	I did not pay attention to differences between the help elements and the real device.
Q27	I saw the virtual elements as real as the device.
Q28	I had the feeling that I could touch the the elements appearing on the device.
Q29	I didn't need to work a lot for recognizing the help elements as 3D elements.
Q30	I liked how the help elements were displayed.
Q31	I liked how the virtual elements were integrated with the real device.
Q32	I think I have learnt with this application.
Q33	I would like to use this technology for other uses.
Q34	I liked this experience.
Q35	I have been concentrated on the tasks to be done, not in the smartphone.
Q36	I have felt involved in the experience.
Q37	I have felt expert on the application at the end of the experience.
Q38	Assess the 3D.
Q39	Mark the application.
Q40	Mark the utility of the application as a help for learning.

the same functionality explained in the written instructions, in order to improve the understanding of the working and operation of both alternatives. Then, a staff person checked that the participant understood and completely knew both applications (AR app and video browser app), giving some additional explanations if required. At that point, the participant filled out a questionnaire with his/her personal and professional demographics.

After this initial training, the final purpose of the experiment was explained to the participants. The purpose was the replacement of some components of a general purpose consumer device, a small size PC widely used at industrial levels in office IT environments. The replacement of the components had to be conducted in the shortest time and making the least number of mistakes as possible. Participants were instructed not to stop nor delay the task due to any reason beyond the experiment, such as talking about the experiment with any person in the room, or using the smartphone for other purposes. The staff verified that all the participants had switched off their smartphones. No incidents were recorded during the experiments, and no samples had to be rejected.

The participants had to complete two procedures in order, first a simple one and then a more complex one. Once the participants were ready to start the components substitution task, they were provided with a smartphone whose desktop exclusively contained the icons corresponding to the AR and the video browser apps. The mobile video browser app showed two folders, labeled as 'simple procedure' and 'complex procedure', containing the respective demonstrative videos for the tasks. The access to the complex video was secured through a password that was given to the participants only after completing the first procedure. Similarly, the AR app indicated that the user should first complete the simple procedure. These details were also explained to participants in the initial training phase. Participants could perform the tests in an unassisted way. In fact, no interventions were required during the whole experiment. Figure 6 shows a snapshot of a real experiment trial. This figure shows the moment when the user is watching a video.

Once the participants completed their assembly, they should make a triple-tap gesture in the smartphone, which registered the elapsed time assembly. It also notified the staff



Fig. 6 A real participant in the experiment watching a video-tutorial

(through an acoustic signal) to check the assembly and mark its quality. The assembly evaluation ranged in a scale from 10 to 1, discounting 1.5, 1.0, or 0.5 points for each serious, moderate, or minor mistakes done, respectively. After completing each one of the alternatives included in the experiment (AR or VT) participants answered a written questionnaire regarding their experience. The staff added on the questionnaire the time required for each assembly, as well their final evaluation of the assembly quality.

Since the mini tower PC was a very common model in our university, we invited professors, administrative staff, and students of the Science Campus at the University of Valencia (Spain) to perform the experiments. The experiments sessions were organized through Eventbrite [3], a self-service event management and promotion website.

As described above, the participants in the study were split into two groups, denoted as A and B. The reason behind this separation is to check if the order in which the learning methods are used has an effect on both the knowledge acquired or the perception of the APP analyzed. The procedure followed by the participants is illustrated in Fig. 7. Group A participants first watch the online video and they perform the practical tests (Assembly 1). Next, they use the App and again perform the practical tests (Assembly 2). The procedure followed by Group B participants is the opposite: they use the App before performing the practical test (Assembly 1), and then they watch the online video and repeat the tests (Assembly 2).

Concretely, the protocol followed by the participants is the following one: first, all participants complete the PreTest. Then, the participants use the first learning method while performing the Assembly 1 tests. Meanwhile, the staff person fills the Assembly 1 observation test. In this stage, Group A participants watch the online video, and Group B participants use the App including AR. At that point, all participants complete the PostTest1. Next, all participants use the other learning method while performing Assembly 2 tests. Meanwhile, the staff person fills the Assembly 2 observation test. In this stage, Group A participants use the App including AR, and Group B participants watch the online video. Next, all participants complete the PostTest2. Finally, all participants complete the Usability and Satisfaction Test.

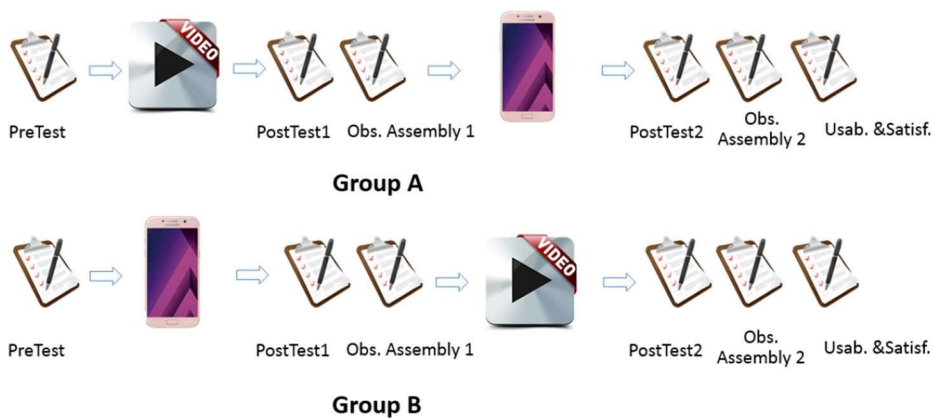


Fig. 7 a Procedure followed by Group A. b Procedure followed by Group B

5 Results

In this Section, we analyze the data obtained about the participants learning as well as the participant satisfaction with both the mobile AR app and the training based on video-tutorials. All the data have been collected using questionnaires. We have used the statistics open-source program R (<http://www.r-project.org>). For all of the results shown below, all significance tests were two-tailed and conducted at the 0.05 significance level. Also, we have included in the results metrics related to the effect size of the performed experiments, in terms of Cohen's d and Eta-squared (η^2). These type of metrics have yielded valuable information about the size of the samples when comparing the results of our experiments. This property is not evaluated by the p -values. All the underlined values in the tables shown in this section mean statistically significant values.

5.1 Learning factor

We analyzed the data in questionnaires PreTest, PostTest1, and PostTest2 in order to evaluate the participants' learning when using the VT and AR mobile app. For each of the questionnaires, we created a 'knowledge' variable summarizing the answers given for each test. First, we analyzed if the collected data follow a normal distribution. Although they are not shown here for the sake of shortness, the Kolmogorov-Smirnov tests [26] ($D = 0.23271$, p -value = 0.02627), Anderson-Darling tests [2] ($A = 2.2765$, p -value = 7.035e-06), as well as Shapiro-Wilk tests [47] ($W = 0.87389$, p -value = 0.0003616) indicated that the learning data do not follow a normal distribution, and therefore we used non-parametric tests to analyze the collected data. In particular, the Wilcoxon Signed-rank sum test (for paired data) and the Mann-Whitney test (for unpaired data) [36] were performed to observe the effect of the methods in the knowledge scores for the entire experiment. The results showed that the global scores for both in the PostTest1 (5.47 ± 1.84) [$W = 29.5$, $p < 0.001$] and the PostTest2 (5.68 ± 0.89) [$W = 7.5$, $p < 0.001$] were significantly higher than the scores in the PreTest (5.11 ± 0.73), indicating the use of these information systems improve the learning experienced by the users, in general terms. Since the scores in the PostTest2 are also significantly higher than the scores in the PostTest1 [$W = 39$, $p < 0.001$], we can state that the use of a second training method (training users using the AR tool, once they have been trained using video tutorials or vice versa) adds a significant improvement in their learning process. However, a much more exhaustive analysis of the tests results is required (comparing the PreTest, PostTest1, and PostTest2 results of group A to the ones of group B) to study the effectiveness of each learning method (VT or AR mobile app).

For this purpose, we have performed the Signed-rank test (for paired data) on the results obtained from each of the groups. Concretely, these tests measure the level of knowledge of the participants in group A on PC maintenance operations before starting the tests (dataset labeled as PreTest), and we have compared these results to the level of knowledge acquired with the training based on video tutorials (dataset labeled as PostTest1 A) and also to the next training stage where the AR tool was used (dataset labeled as PostTest2 A). In the same way, we have compared the level of knowledge of the participants in group B before the training (dataset labeled as PreTest B) to the level of knowledge after the training with the AR tool (dataset labeled as PostTest1 B) and the use of video tutorials (dataset labeled as PostTest2 B).

Additionally, we have performed the Wilcoxon Mann-Whitney U test (for unpaired data) to compare the knowledge results between different groups during the same training stage.

Table 3 Mann-Whitney U test (MW) and Wilcoxon Signed-rank (WS) tests results comparing the learning of the participants

Comparison		Test	U/W	Z	p	r
Pre A	Pre B	MW	U=216	0.434	0.674	0.069
Pre A	Post1 A	WS	W=9.5	-3.356	<0.002	0.531
Pre A	Post2 A	WS	W=3	-3.757	<0.001	0.594
Post1 A	Post2 A	WS	W=3.5	-3.581	<0.001	0.566
Pre B	Post1 B	WS	W=5	-3.644	<0.001	0.576
Pre B	Post2 B	WS	W=1	-3.885	<0.001	0.614
Post1 B	Post2 B	WS	W=18	-2.899	<0.003	0.458
Post1 A	Post1 B	MW	U=205	0.135	0.903	0.021
Post2 A	Post2 B	MW	U=208	0.203	0.850	0.032

This test allows the comparison of the level of knowledge acquired in the first training stage by group A participants (based on VT and labeled as PostTest1 A) to the level of knowledge acquired in the same stage by group B participants (based on AR and labeled PostTest1 B).

Table 3 shows the learning results using both tests, as well as the significance level, for all the datasets for both participants groups. The underlined values means statistically significant differences. All the labels referencing PreTests and PostTests have been shortened as 'Pre' and 'Post', respectively, due to space reasons.

The first row in Table 3 shows that the starting level of knowledge of both groups (A and B) is similar, since there are no significant differences in the average values (U= 216, p = 0.674). This table also shows that both groups significantly improved their knowledge when using the first learning method, since the second and fifth row shows statistically significant differences (W = 9.5, p < 0.002, and W = 5, p < 0.002). However, the level of knowledge acquired by both groups can be considered as similar, since there is no a significant difference when comparing PostTest1 A with PostTest1 B results (U= 205, p = 0.903 in the second last row). Similar results are obtained for the second training stage with a second learning tool. Both groups significantly improve their knowledge. However, the last row in the table shows that there is not a significant difference in the knowledge acquired after having used both learning tools (U= 208, p = 0.850). Therefore, we can conclude that both methods can be considered as valid for user's training in minor maintenance operations for conventional consumer electronic devices.

Table 4 Kruskal-Wallis test results for evaluating the learning of the participants depending on different factors

Factor	PreTest		PostTest1		PostTest2	
	KW(x2)	p-val.	KW(x2)	p-val.	KW(x2)	p-val.
Gen.	10.217	<0.002	8.261	<0.005	8.954	<0.003
Age	22.051	0.077	22.544	0.068	22.325	0.0721
CS	0.003	0.956	0.003	0.956	0.0030	0.956
PE	10.217	<0.002	15.446	<0.001	15.051	<0.001

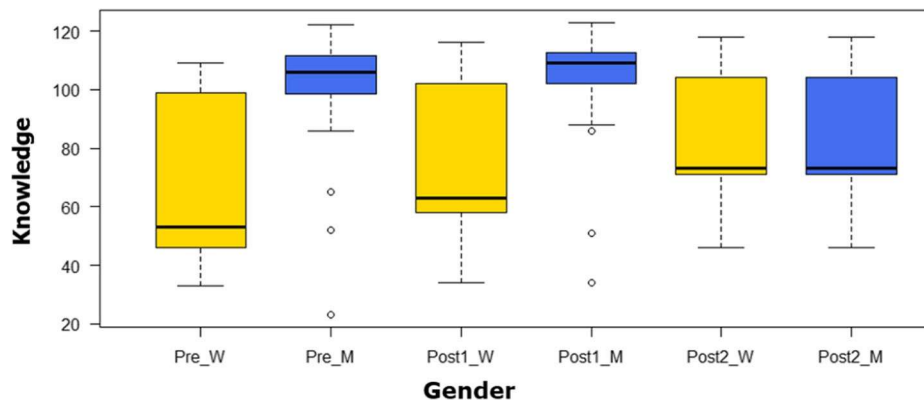


Fig. 8 Boxplot representation of the global knowledge by gender

A Kruskal-Wallis test [11] was also performed in order to take into consideration several factors simultaneously. The factors of training system such as gender, age, professional experience and computer skills between subjects were considered, and the results are shown in Table 4. The labels ‘Gen.’, ‘CS’ and ‘PE’ respectively stand for ‘Gender’, ‘Practical Computer Science Skills’ and ‘Professional Experience’. The underlined values indicate statistically significant differences. The first and last rows in this table shows that practical computer science skills and gender had a statistically significant effect on the learning factor.

Figures 8 and 9 graphically show these results using a boxplot representation. Figure 8 shows that the different starting level of knowledge about maintenance operations for conventional consumer electronic devices between gender was significant. However, this difference considerably decreased as they carried out the training sessions. It seems likely that a third training stage could remove the statistically significant difference between genders.

Figure 9 shows the level of knowledge of the participants grouped by their practical computer science skills. For this study, we classified the participants into three categories: professionals/students with basic office IT skills (labeled as T1), IT professionals (labeled as T2), and students of a computer science related degree (labeled as T3). This figure shows

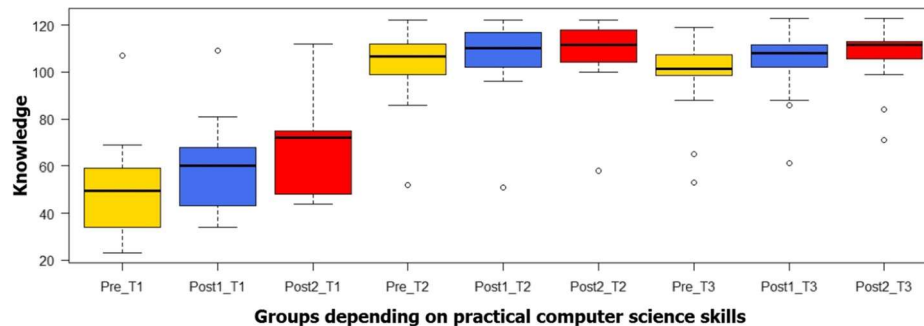


Fig. 9 Boxplot representation of the global knowledge by computer science skills

Table 5 Assembly observation test results

Id	Question	Test 1	Test 2
OE1	Assembly 1 quality	7±0.75	7±0
OT1	Time for Assem. 1 compl.	5.3±2.1	3.4±1.63
OE2	Assembly 2 quality	7±0.7	7±1
OT2	Time for Assem. 2 compl.	13.05±2.55	10.65±3.3

that the initial knowledge of each group is different and the training stages help all groups to gain knowledge.

5.2 Quality and time for task completion

We also analyzed the results of the observation tests for both assemblies (Assembly observation questionnaires). Each staff person measured two metrics: the time required for the assembly completion (variables OT1 and OT2), and the quality of the final assembly (variables OE1 and OE2). The results for these variables are shown in Table 5. OE1 and OE2 values ranged from 1 to 7, being 7 the highest quality. The 'task completion time' was measured and reported in minutes. This table shows that the quality of the assemblies was high, with a median of 7 for both tests (OE1 and OE2). It must be noted that the assembly was performed while using the corresponding learning tool (video tutorial or mobile AR app).

We performed a more detailed analysis of these data. Concretely, a Mann-Whitney test (for unpaired data) and a Wilcoxon Signed-rank sum (for paired data) were performed to observe the effects of the learning method used on both the time required for completion and the assembly quality. Table 6 shows the results for the Mann-Whitney U test (MW) and Wilcoxon Signed-rank (WS) comparing OT1 and OT2 variables. First, we analyze the data for all the participants (PostTest1 (PT1) and PostTest2 (PT2)), and next we analyze the participants separated by groups (PostTest1 Group A (PT1A), PostTest2 Group A (PT2A), PostTest1 Group B (PT1B), and PostTest2 Group B (PT2B)). The underlined values indicate statistically significant differences.

Table 6 shows that there are significant differences in the required time when performing the assembly for the first time, in regard to the second time (PostTest1 and PostTest2),

Table 6 Mann-Whitney U test (MW) and Wilcoxon Signed-rank (WS) for the completion times

Data	Assembly 1		Assembly 2	
	U/W	p	U/W	p
PT1-PT2	W=505.5	< <u>0.001</u>	W=519.5	< <u>0.001</u>
PT1A-PT2A	W=121	<u>0.007</u>	W=147.5	< <u>0.001</u>
PT1B-PT2B	W=137	<u>0.005</u>	W=126	<u>0.017</u>
PT1A-PT1B	U=93	0.078	U=84.5	<u>0.040</u>
PT2A-PT2B	U=141.5	0.931	U=122.5	0.459

regardless of the order in which the learning methods are used. There are significant differences in the time required for both groups, i.e. longer times are required by group B when using the App. However, there are no significant differences in the time required by both groups to perform the second assembly. The reason may be that the assembly was more complex, and it required group B participants to watch the whole video. A curious detail is that the time values consumed in the assembly when using the App are longer than then the ones consumed when using the video. The reason, according to the observation staff, is that when the participants used the App they like the AR so much that they wanted to enjoy and watch it from different angles, delaying the task completion. However, they forwarded some parts of the video tutorial. The opinion of the observation staff was that the App was much more attractive and stimulating to the participants than the video tutorial.

Table 7 shows the comparison of the quality variables, OE1 and OE2. This table first shows the data for all the participants (PreTest (PR), PostTest1 (PT1), and PostTest2 (PT2)). Next, it shows the data for the participants separated by groups (PreTest Group A (PRA), PostTest1 Group A (PT1A), PostTest2 Group A (PT2A), PreTest Group B (PRB), PostTest1 Group B (PT1B), and PostTest2 Group B (PT2B)).

Table 7 shows that the assembly quality improved after any of the learning methods, and it also improved after using both methods. However the order in which the methods were used did not affect the results. Therefore, we can conclude that both learning methods help the participants to learn how to perform the assemblies in a similar way, significantly contributing to the learning process.

5.3 Satisfaction and usability

Presence can be defined in virtual environments as an individual and context-dependent user response, related to the experience of ‘being there’ [9]. According to Regenbrencht & Schubert [42], this definition cannot be exactly applied to AR. However, presence can also

Table 7 Mann-Whitney U test (MW) and Wilcoxon Signed-rank (WS) for the quality variables

Data	Assembly 1		Assembly 2	
	U/W	p	U/W	p
PR-PT1	W=0	<0.001	W=0	<0.001
PR-PT2	W=0	<0.001	W=0	<0.001
PT1-PT2	W=0	0.021	W=25	0.492
PRA-PT1A	W=0	<0.001	W=0	<0.001
PRA-PT2A	W=0	<0.001	W=0	<0.001
PT1A-PT2A	W=0	0.174	W=3	0.233
PRB-PT1B	W=0	<0.001	W=0	<0.001
PRB-PT2B	W=0	<0.001	W=0	<0.001
PT1B-PT2B	W=0	0.098	W=8.5	0.751
PT1A-PT1B	U=156.5	0.610	U=151.5	0.781
PT2A-PT2B	U=153	0.653	U=161	0.476

be achieved in AR, measuring the experienced presence of virtual elements in the real environment [42]. In order to measuring this AR sense of presence, we have designed the Q2 questionnaire. The questionnaires of Wittmer & Singer's [54] and Regenbrech & Schubert [42] were used as the basis for (and they were adapted to) our study, as we did in many other investigations [16, 25, 44]. The usability and satisfaction questionnaire is shown in Table 2, containing forty questions about six different factors [54]: Eight questions related to Control Factors (CF): Degree of control, immediacy of control, anticipation of events, mode of control, physical environment modifiability; Four questions related to Sensory Factors (SF): Sensory modality, environmental richness, multimodal presentation, consistency of multimodal information, degree of movement perception, active search; Two questions related to Distraction Factors (DF): selective attention, interface awareness; Eleven questions related to Realism Factors (RF): scene realism, information consistent with objective world, meaningfulness of experience. We have also added seven questions related to ergonomics (EF): devices comfort when using, effort, etc., and nine other questions (OF), scoring different aspects related to the experiment: 3D perception, usefulness, etc.

We analyzed the usability and satisfaction of the participants with different aspects of the developed mobile AR App, starting from the answers in the usability and satisfaction questionnaire. The Kolmogorov-Smirnov test, ($D = 0.10129$, $p\text{-value} = 0.8185$), Anderson-Darling test ($A = 0.45265$, $p\text{-value} = 0.2582$), and Shapiro-Wilk test ($W = 0.96476$, $p\text{-value} = 0.256$) indicated that data follow a normal distribution, and therefore we have used parametric tests (the t-test and the Cohen's test for paired data and the ANOVA test). The answers to the forty questions showed that the user satisfaction with the App was high, being above 4 in all the questions. Question OF8 asked the user to rate the App (from 1 to 7), and the average and standard deviation for this question was 6.05 ± 0.93 , showing a high degree of satisfaction. Participants are also of the opinion that the App is very useful in their learning process (question OF9).

Table 8 shows the Cohen's test for the gender and professional factors, including both the general satisfaction with the App and the satisfaction grouped by types of questions. This table shows that there are no significant differences, with two exceptions. First, the fourth row in the table shows a statistically significant difference ($t(37)=-2.23$, $p= 0.032$, Cohen's $d= 0.78$) between the average value of the control factors (CF) for men (40.8 ± 7.41) after using the mobile AR app and the value for women (34.6 ± 8.53). Second, the fourth row shows that the average value for the distraction factors (DF) is significantly higher ($t(37)=-2.43$, $p= 0.020$, Cohen's $d= 0.84$) for men (10.30 ± 2.03) when compared to the one for

Table 8 Cohen's test for the gender and professional factors

Satisf.	Gender			Student/Professional		
	t	p-value	Cohen's d	t	p-value	Cohen's d
Global	-1.58	0.136	0.53	0.09	0.926	0.03
SF	0.54	0.596	0.19	-1.10	0.279	0.35
EF	-0.02	0.983	0.01	0.65	0.520	0.21
CF	-2.23	<u>0.032</u>	0.78	0.26	0.795	0.08
DF	-2.43	<u>0.020</u>	0.84	0.65	0.521	0.21
RF	-1.43	0.161	0.50	0.18	0.860	0.06
OF	-1.92	0.062	0.67	0.09	0.926	0.03

Table 9 Multifactorial ANOVA test about global satisfaction and satisfaction by groups depending on different factors

Var.	Gender			Student/Professional		
	F-value	p-value	η^2	F-value	p-value	η^2
Global	1.397	0.254	0.447	0.758	0.396	0.0173
SF	1.826	0.119	0.552	0.004	0.948	$9*10^{-5}$
EF	0.792	0.667	0.328	0.576	0.458	0.017
CF	1.314	0.293	0.455	0.760	0.396	0.019
DF	0.707	0.741	0.287	2.640	0.123	0.0765
RF	1.444	0.234	0.440	0.835	0.374	0.018
OF	1.302	0.299	0.380	6.015	0.0253	0.125

women (8.42 \pm .47). In all the cases shown in this table, the degree of freedom (df) value was 37.

Table 9 shows the participants' satisfaction depending on different factors like age and type of participants. The results in this table show that the user satisfaction does not depend on the age nor the kind of initial knowledge of the user, taking into account the significance level of the experiment ($p < 0.05$).

We checked next if there was a correlation between the learning of the participants and the satisfaction with the App. Table 10 shows the Pearson correlation among the satisfaction with the App, the previous learning, and the learning after using each of the learning methods. The results in this table shows positive correlation between the satisfaction and learning variables. It also shows a positive but lower correlation between satisfaction and the observation marks. These results suggest that the learning acquired by the participants affects the level of satisfaction.

We also asked the participants to give their views on the App. Participants consider that the App is more original, attractive, funny, fast, simple, and entertaining than the videos. They think that videos tire them and are outdated. Many users have highlighted that the App seemed a game. However, one of the most interesting comments is that users think that they wasted less time and they made less mistakes when performing the tests with the App, because they directly watched what to do at each moment on the PC. They also think that the App induced them towards a step-by-step procedure, without skipping any step. Thus, they think it is difficult to make mistakes with the mobile AR App.

Table 10 Pearson correlation among different variables of the study

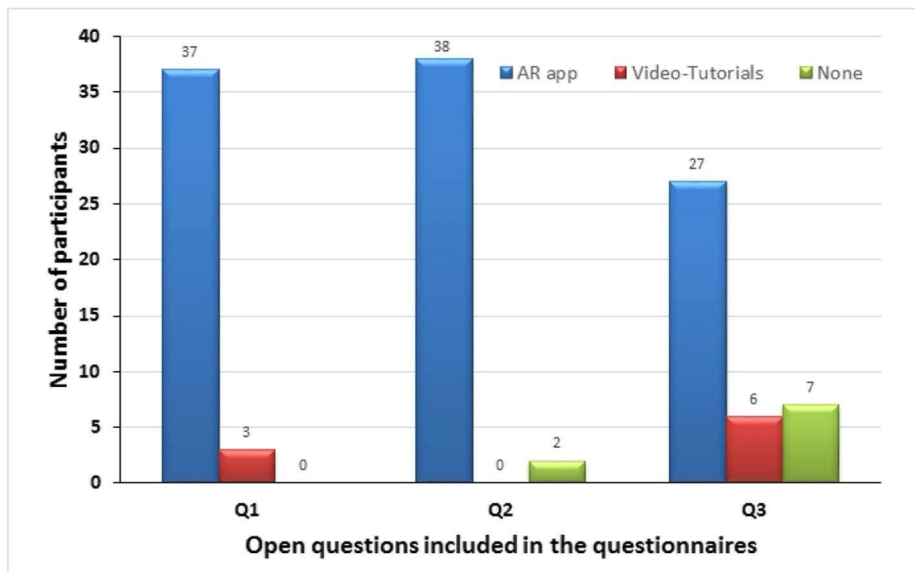
Variable	T	df	p-value	Corr
PreTest learn.	5.2584	37	<u><0.000007</u>	0.6539843
PostTest1 learn.	5.1205	37	<u><0.00001</u>	0.6440015
PostTest2 learn.	5.421	37	<u><0.000004</u>	0.6653274
Assem. 1 observ. – test 1	3.8857	35	<u><0.0005</u>	0.5489775
Assem. 1 observ. – test 2	2.1775	33	<u>0.0367</u>	0.3544411
Assem. 2 observ. – test 1	2.1963	35	<u>0.03479</u>	0.3480275
Assem. 2 observ. – test 2	2.4893	35	<u>0.0177</u>	0.3878385

Table 11 Open questions where users preferred the App

Question	Mean±sd	Median±IQ	Answer
More helpful method	1.08±0.27	1±0	App
Method you like the most	0.95±0.22	1±0	App
Method you would recommend	0.97±0.58	1±0	App

We included three questions in the Usability and satisfaction questionnaire about the user preferences between the two learning methods. Since these questions allowed open answers, most of the users answered ‘App’ or ‘Video’, although some of them answered ‘none’. The results for the answers choosing the App are shown in Table 11.

Finally, Fig. 10 shows the absolute numbers about the answers collected for the questions in Table 11. We have denoted as ‘Q1’, ‘Q2’, and ‘Q3’ the questions in the first, second and third row of that table, respectively. This figure shows that most of the participants prefer the App which they liked the most. However, the answers to the third question shows that 27 of the 40 participants (67%) would recommend the App, 6 participants (15%) would recommend the video, and 7 participants (18%) would recommend none of them. When analyzing the rest of the answers (not shown here) these results are explained: the users recommending the video learning method think that their company will never purchase or develop Apps like this, and they probably will create videos instead. The users who do not like any of the methods prefer more conventional learning methods, like human teachers or paper-based tutorials.

**Fig. 10** Answers to questions in Table 11

6 Conclusion

The worldwide need for environmental sustainable processes has led consumer electronics industry to replace the traditional, paper-based operation and maintenance user guides included in consumer devices by multimedia tutorials supported on electronic media. More recently, some consumer electronics companies have started to offer standard and pre-packaged AR solutions as alternative to video-based tutorials for minor maintenance operations.

In this work, we have presented the results of a study carried out to determine if mobile AR apps are more effective than traditional video tutorials in helping users learn about minor maintenance operations on consumer electronic devices. The results indicate that both systems help users to complete the maintenance operations properly while achieving knowledge in various aspects of equipment maintenance. No significant differences appeared between the knowledge acquired when using each of the learning methods, corroborating our primary hypothesis. Although no statistically significant differences were found between AR and VT solutions, users scored higher on the AR version in all cases. Therefore, these results do not formally validate our secondary hypothesis. However, users explicitly preferred the AR version when following three different usability and satisfaction criteria, validating our third hypothesis. For the AR version, a strong and significant correlation was found between the satisfaction and the achieved knowledge. Since the AR solution achieved similar learning results with higher usability and satisfaction scores than the video-based tutorials, these results suggest that AR solutions are the most effective approach to substitute the typical paper-based instructions in consumer electronic devices.

References

1. Ahn J, Williamson J, Gartrell M, Han R, Lv Q, Mishra S (2015) Supporting healthy grocery shopping via mobile augmented reality. *ACM Trans Multimedia Comput Commun Appl* 12(1s):16:1–16:24. <https://doi.org/10.1145/2808207>
2. Anderson TW (2011) Anderson–darling tests of goodness-of-fit. Springer, Berlin, pp 52–54. https://doi.org/10.1007/978-3-642-04898-2_118
3. Awad N, Lewandowski SE, Decker EW (2015) Event management system for facilitating user interactions at a venue. US Patent App. 14/829,382
4. Azuma R (1997) A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6(4):355–385
5. Baird K, Barfield W (1999) Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality* 4:250–259
6. Ballo P (2018) Hardware and software for ar/vr development. In: *Augmented and virtual reality in libraries*, pp 45–55. LITA guides
7. Barrile V, Fotia A, Bilotta G (2018) Geomatics and augmented reality experiments for the cultural heritage. *Applied Geomatics*. <https://doi.org/10.1007/s12518-018-0231-5>
8. Billingham M, Duenser A (2012) Augmented reality in the classroom. *Computer* 45(7):56–63. <https://doi.org/10.1109/MC.2012.111>
9. Bowman DA, McMahan RP (2007) Virtual reality: how much immersion is enough? *Computer* 40(7)
10. Brown TA (2015) *Confirmatory factor analysis for applied research*. Guilford Publications
11. Dodge Y.(ed) (2008) *Kruskal-Wallis test*. Springer, New York. https://doi.org/10.1007/978-0-387-32833-1_216
12. Elmunsyah H, Hidayat WN, Asfani K (2019) Interactive learning media innovation: utilization of augmented reality and pop-up book to improve user's learning autonomy. *J Phys Conf Ser* 1193(012):031. <https://doi.org/10.1088/1742-6596/1193/1/012031>
13. Entertainment L (2017) *Dolphin Player*. <https://play.google.com/store/apps/details?id=com.broov.player>. Online; Accessed 09-September-2017

14. Fletcher J, Belanich J, Moses F, Fehr A, Moss J (2017) Effectiveness of augmented reality & augmented virtuality. In: MODSIM Modeling & simulation of systems and applications) world conference
15. Fraga-Lamas P, Fernández-Caramés TM, Blanco-Novoa O, Vilar-Montesinos MA (2018) A review on industrial augmented reality systems for the industry 4.0 shipyard. *IEEE Access* 6:13,358–13,375. <https://doi.org/10.1109/ACCESS.2018.2808326>
16. Furió D, Juan MC, Seguí I, Vivó R (2015) Mobile learning vs. traditional classroom lessons: a comparative study. *J Comput Assist Learn* 31(3):189–201. <https://doi.org/10.1111/jcal.12071>
17. Gavish N, Gutiérrez T, Webel S, Rodríguez J, Peveri M, Bockholt U, Tecchia F (2015) Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interact Learn Environ* 23(6):778–798. <https://doi.org/10.1080/10494820.2013.815221>
18. Gimeno J, Morillo P, Orduña JM, Fernández M (2013) A new ar authoring tool using depth maps for industrial procedures. *Comput Ind* 64(9):1263–1271. <https://doi.org/10.1016/j.compind.2013.06.012>
19. Holzinger A, Kickmeier-Rust MD, Albert D (2008) Dynamic media in computer science education; content complexity and learning performance: is less more? *Educational Technology & Society* 11(1):279–290
20. Hornbaek K (2013) Some whys and hows of experiments in human–computer interaction. *Foundations and Trends® in Human–Computer Interaction* 5(4):299–373. <https://doi.org/10.1561/11000000043>
21. Huang J, Liu S, Xing J, Mei T, Yan S (2014) Circle & search: Attribute-aware shoe retrieval. *ACM Trans Multimedia Comput Commun Appl* 11(1):3:1–3:21. <https://doi.org/10.1145/2632165>
22. Jiang S, Wu Y, Fu Y (2018) Deep bidirectional cross-triplet embedding for online clothing shopping. *ACM Trans Multimedia Comput Commun Appl* 14(1):5:1–5:22. <https://doi.org/10.1145/3152114>
23. Kim SK, Kang SJ, Choi YJ, Choi MH, Hong M (2017) Augmented-reality survey: from concept to application. *KSII Transactions on Internet and Information Systems* 11:982–1004. <https://doi.org/10.3837/tiis.2017.02.019>
24. Langlotz T, Zingerle M, Grasset R, Kaufmann H, Reitmayr G (2012) Ar record&replay: Situated compositing of video content in mobile augmented reality. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, OzCHI '12. ACM, New York, pp 318–326. <https://doi.org/10.1145/2414536.2414588>
25. Martín-SanJose JF, Juan MC, Mollá R, Vivó R (2017) Advanced displays and natural user interfaces to support learning. *Interact Learn Environ* 25(1):17–34. <https://doi.org/10.1080/10494820.2015.1090455>
26. Massey FJ (1951) The kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 46(253):68–78
27. van der Meij H, van der Meij J, Voerman T, Duipmans E (2018) Supporting motivation, task performance and retention in video tutorials for software training. *Educ Technol Res Dev* 66(3):597–614. <https://doi.org/10.1007/s11423-017-9560-z>
28. van der Meij J, van der Meij H (2015) A test of the design of a video tutorial for software training. *J Comput Assist Learn* 31(2):116–132. <https://doi.org/10.1111/jcal.12082>
29. Mestre LS (2012) Student preference for tutorial design: a usability study. *Ref Serv Rev* 40(2):258–276. <https://doi.org/10.1108/00907321211228318>
30. Mohr P, Kerbl B, Donoser M, Schmalstieg D, Kalkofen D (2015) Retargeting technical documentation to augmented reality. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, CHI '15. ACM, New York, pp 3337–3346. <https://doi.org/10.1145/2702123.2702490>
31. Mohr P, Mandl D, Tatzgern M, Veas E, Schmalstieg D, Kalkofen D (2017) Retargeting video tutorials showing tools with surface contact to augmented reality. In: Proceedings of the 2017 CHI conference on human factors in computing systems, CHI '17. ACM, New York, pp 6547–6558. <https://doi.org/10.1145/3025453.3025688>
32. Montgomery DC, Runger GC (2003) Applied statistics and probability for engineers. Wiley, New York
33. Morillo P, Orduña JM, Casas S, Fernández M (2019) A comparison study of ar applications versus pseudo-holographic systems as virtual exhibitors for luxury watch retail stores. *Multimedia Systems*. <https://doi.org/10.1007/s00530-019-00606-y>
34. Morse JM (2000) Determining sample size. *Qual Health Res* 10(1):3–5. <https://doi.org/10.1177/104973200129118183>
35. Muñoz-Montoya F, Juan M, Mendez-Lopez M, Fidalgo C (2019) Augmented reality based on slam to assess spatial short-term memory. *IEEE Access* 7:2453–2466. <https://doi.org/10.1109/ACCESS.2018.2886627>
36. Neuhäuser M (2011) Wilcoxon–Mann–Whitney test. Springer, Berlin, pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615
37. Neumann U, Majoros A (1998) Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In: Inproceedings of the IEEE virtual reality annual international symposium (VR '98), pp 4–11

38. no JJA, Juan MC, Gil-Gómez JA, Mollá R. (2014) A comparative study using an autostereoscopic display with augmented and virtual reality. *Behaviour & Information Technology* 33(6):646–655. <https://doi.org/10.1080/0144929X.2013.815277>
39. Palmarini R, Erkoyuncu JA, Roy R, Torabmostaedi H (2018) A systematic review of augmented reality applications in maintenance. *Robot Comput Integr Manuf* 49:215–228
40. Quint F, Loch F (2015) Using smart glasses to document maintenance processes. *Mensch und Computer 2015–Workshopband*
41. Radkowski R, Herrema J, Oliver J (2015) Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *International Journal of Human–Computer Interaction* 31(5):337–349. <https://doi.org/10.1080/10447318.2014.994194>
42. Regenbrecht H, Schubert T (2002) Measuring presence in augmented reality environments: design and a first test of a questionnaire, Porto, Portugal
43. Robertson J (2012) Likert-type scales, statistical methods, and effect sizes. *Commun ACM* 55(5):6–7. <https://doi.org/10.1145/2160718.2160721>
44. Rodríguez-Andrés D, Juan MC, Méndez-López M, Pérez-Hernández E, Lluch J (2016) Mnemocity task: Assessment of childrens spatial memory using stereoscopy and virtual environments. *PLoS ONE* 11(8). <https://doi.org/10.1371/journal.pone.0161858>
45. Sanna A, Manuri F, Lamberti F, Paravati G, Pezzolla P (2015) Using handheld devices to support augmented reality-based maintenance and assembly tasks. In: 2015 IEEE International conference on consumer electronics (ICCE), pp. 178–179. <https://doi.org/10.1109/ICCE.2015.7066370>
46. Schmidt S, Ehrenbrink P, Weiss B, Voigt-Antons J, Kojic T, Johnston A, Moller S (2018) Impact of virtual environments on motivation and engagement during exergames. In: 2018 Tenth international conference on quality of multimedia experience (qoMEX), pp 1–6. <https://doi.org/10.1109/QoMEX.2018.8463389>
47. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52(3/4):591–611
48. Tang A, Owen C, Biocca F, Mou W (2003) Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '03. ACM, New York, pp 73–80, <https://doi.org/10.1145/642611.642626>
49. Tomás JM, Oliver A, Galiana L, Sancho P, Lila M (2013) Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling: A Multidisciplinary Journal* 20(2):299–313. <https://doi.org/10.1080/10705511.2013.769394>
50. Uva AE, Gattullo M, Manghisi VM, Spagnulo D, Cascella GL, Fiorentino M (2017) Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *The Int J Adv Manuf Technol*: 1–13
51. Wang X, Ong SK, Nee AYC (2016) A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4(1):1–22. <https://doi.org/10.1007/s40436-015-0131-4>
52. Westerfield G, Mitrovic A, Billingham M (2015) Intelligent augmented reality training for motherboard assembly. *Int J Artif Intell Educ* 25(1):157–172. <https://doi.org/10.1007/s40593-014-0032-x>
53. Wiedenmaier S, Oehme O, Schmidt L, Luczak H (2003) Augmented reality (ar) for assembly processes - design and experimental evaluation. *International Journal of Human-Computer Interaction* 16(3):497–514
54. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7(3):225–240
55. Wu HK, Lee SWY, Chang HY, Liang JC (2013) Current status, opportunities and challenges of augmented reality in education. *Computers & Education* 62:41–49. <https://doi.org/10.1016/j.compedu.2012.10.024>
56. Yim MYC, Chu SC, Sauer PL (2017) Is augmented reality technology an effective tool for e-commerce? an interactivity and vividness perspective. *Journal of Interactive Marketing* 39(<http://www.sciencedirect.com/science/article/pii/S1094996817300336>):89–103. <https://doi.org/10.1016/j.intmar.2017.04.001>
57. Yuan ML, Ong SK, Nee AYC (2008) Augmented reality for assembly guidance using a virtual interactive tool. *Int J Prod Res* 46(7):1745–1767. <https://doi.org/10.1080/00207540600972935>



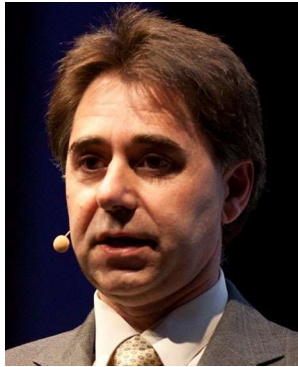
Pedro Morillo received the MS and PhD degrees in computer engineering from the University of Valencia, Spain, with a dissertation on “Improving the Performance in Distributed Virtual Environments.” Currently, he is an associate professor in the Department of Informatics, University of Valencia. In this department, he belongs to the Networking and Virtual Environments (GREV) group, where he focuses on the design and development of network architectures for distributed virtual environments. Furthermore, his research interests include distributed virtual environment systems, load balancing, metaheuristics, and cluster computing. He also served as a visiting scientist at Iowa State University, Ames, Iowa, and the University of Louisiana, Lafayette (USA).



Inmaculada García-García received the MS degree from the Technical University of Valencia and the PhD degree in computer engineering from the University of Valencia, Spain, with a dissertation on “Hybrid simulation as real time graphic applications kernel”. Currently, she is an associate profesor in the Department of Informatics Systems and Computation, Technical University of Valencia. She belongs to the Valencian Research Institute for Artificial Intelligence (VRAIN), where she focuses on the design on intelligent services, more specifically in the recommender systems design.



Juan M. Orduña received the MS in computer engineering from the Technical University of Valencia, Spain, in 1990. He worked at Telefónica de España, Manpel Electrónica, S.A. and at the Technical University of Valencia as a computer engineer. He received the PhD. in computer engineering from the University of Valencia in 1998. His research has been developed inside the ACCA team (<http://www.acca-group.info>). Currently, he is a full professor in the Department of Informatics, at the University of Valencia, SPAIN, where he leads the GREV research group (<http://grev.uv.es>). He is member of the HiPEAC network of excellence (<http://www.hipeac.net/>), and his research is currently supported by the Spanish MEC and the European Commission through several projects. Dr. Orduna served as a member of the Program Committee in different conferences and workshops (ICPP, Europar, VR, ICPADS, etc.) as well as a reviewer for scholarly journals like Parallel Computing, IEEE TPDS, JNCA, or FGCS. He has published a number of research works in different high impact international journals and conferences.



Marcos Fernández received the M.Sc. degree in physics and the Ph.D. degree in computer science from the University of Valencia. His Ph.D. degree has been oriented to the generation of traffic for driving simulators. He is a lecturer professor in the Department of Informatics at the University of Valencia, Spain. He is also the coordinator of the ARTEC Graphics Group at the Robotics Institute. Currently, his research addresses highly immersive virtual environments. Currently, he is involved in the development of a immersive visualization facility call VISIONARC.



M. Carmen Juan is Full Professor at the UPV. From 1996 she has been lectured at the UPV. She received her M.Sc. Degree in Computer Science in 1994, and from this year she has been working as a researcher in the UPV. In 2000, she received her Ph.D. degree from the UPV. Her research interests have been mainly: Computer Graphics, specifically Augmented Reality, advanced user interfaces and their application to education, medicine and psychology.