Virtual Reality & Intelligent Hardware

• Article •

December 2022 Vol. 4 No. 6: 487–505
10.1016/j.vrih.2022.06.004

Empowering digital twins with eXtended reality collaborations

Lorenzo STACCHIO[1], Alessia ANGELI[2], Gustavo MARFIA[3*]

1. *Department for Life Quality Studies, University of Bologna, Italy;*
2. *Department of Computer Science and Engineering, University of Bologna, Italy;*
3. *Department of the Arts, University of Bologna, Italy*

Received 12 March 2022; **Revised** 5 May 2022; **Accepted** 12 June 2022

Abstract: Background The advancements of Artificial Intelligence, Big Data Analytics, and the Internet of Things paved the path to the emergence and use of Digital Twins (DTs) as technologies to "twin" the life of a physical entity in different fields, ranging from industry to healthcare. At the same time, the advent of eXtended Reality (XR) in industrial and consumer electronics has provided novel paradigms that may be put to good use to visualize and interact with DTs. XR technologies can support human-to-human interactions for training and remote assistance and could transform DTs into collaborative intelligence tools. **Methods** We here present the Human Collaborative Intelligence empowered Digital Twin framework (HCLINT-DT) integrating human annotations (e.g., textual and vocal) to allow the creation of an all-in-one-place resource to preserve such knowledge. This framework could be adopted in many fields, supporting users to learn how to carry out an unknown process or explore others' past experiences. **Results** The assessment of such a framework has involved implementing a DT supporting human annotations, reflected in both the physical world (Augmented Reality) and the virtual one (Virtual Reality). **Conclusions** The outcomes of the interface design assessment confirm the interest in developing HCLINT-DT-based applications. Finally, we evaluated how the proposed framework could be translated into a manufacturing context.

Keywords: Digital twin; eXtended reality; Human collaborative intelligence

Supported by the University of Bologna Alma Attrezzature 2017 grant, AEFFE S.p.a., the Golinelli Foundation and Elettrotecnica Imolese S.U.R.L.

Citation: Lorenzo STACCHIO, Alessia ANGELI, Gustavo MARFIA. Empowering digital twins with eXtended reality collaborations. Virtual Reality & Intelligent Hardware, 2022, 4(6): 487–505

1 Introduction

The advancements of Artificial Intelligence (AI), Big Data Analytics (BDA), and the Internet of Things (IoT) paved the path to the emergence and use of Digital Twins (DTs) in many fields, including manufacturing, aerospace, healthcare, and medicine[1]. DTs are computer models that simulate, emulate, mirror, or "twin" the life of a physical entity, which may be an object, a process, or a human[1]. Unlike simulation models, DTs connect to their physical counterparts supporting bidirectional relationships, receiving real-time data to monitor their operating status, controlling their processes and functions, and returning insights, such as diagnostics and prognostics. For these characteristics, DTs are being applied in different fields, ranging from

*Corresponding author, gustavo.marfia@unibo.it

industry to healthcare[2,3].

Tao et al. explored a novel concept of DT shop-floor, discussing four key components, including physical shop-floor, virtual shop-floor, shop-floor service system, and shop-floor DT data[4]. They aimed at finding a convergence of the manufacturing physical and the virtual worlds to realize smart interconnections, interactions, control, and management.

eXtended Reality (XR) paradigms may add depth to such discussion with the advent of the Metaverse ecosystem and an increased offer of applications[5]. In particular, recent works have already explored the effectiveness of the combination of DTs and Virtual Reality (VR) models in several different application domains[6–8]. Augmented Reality (AR) techniques have also been put to good use to visualize DTs, assisting users in their everyday activities[9–11]. It is worth noticing that most of such works focus on the DTs of humans and objects and their manipulation through VR paradigms, directly influencing the physical world and vice versa. However, a more limited number of works have focused on the role of Human-Machine Interactions (HMIs) in such models.

HMI focuses on finding natural ways to communicate, cooperate, and interact between humans and machines. In this domain, it is worth citing the work of Josifovska et al.[12] who introduced a DT-empowered multi-modal UI framework to adapt assistance systems to different environmental conditions and human workers. In such a proposal, the DT of a human served the purpose of modeling in a systematic and fine-grained way specific human abilities, peculiarities, and preferences. Instead, Ma et al. explored the adoption of DTs in the product lifecycle of an HMI, including its design, manufacturing, and service[13]. Nevertheless, despite the progress made in automation, sensing, and automated learning, today it is widely accepted that "machines cannot fully replace the unique perception and communication skills of humans"[14]. None of such works, to the best of our knowledge, has considered leveraging the potential of XR and DTs technologies to deliver into HMIs a primary component of any system, the collaborative intelligence created and molded by humans while involved in given activities.

The focus of this work is hence to consider the dynamics of workers interacting with physical objects and virtual models from a Collaborative Intelligence (CLINT) perspective. CLINT is not novel to the scientific community, and many different definitions have been provided depending on the specific perspective[14–17].

To define our proposal, we started from the definition of CLINT introduced by[18]: it involves an extensive collaboration of different team members to solve problems. It can provide more information for designing better solutions than any single member could while giving a non-stop real-time learning opportunity. Moreover, such collaboration has the potential of integrating diverse contributions (different members provide different information/knowledge, skill, and experience to a problem resolution) into a platform to produce a creative solution for successfully solving a problem.

Based on those considerations, in this work, we considered the problem of embedding a Human Collaborative Intelligence (HCLINT) experience within a DT. To address such a problem, we present a CLINT empowered DT-based framework for HMIs where the DT of a process can integrate human annotations (e.g., textual and vocal annotations), called HCLINT-DT. The role of this framework is not only to support the exploration of information, but also to allow the creation of an all-in-one-place resource to preserve human knowledge. The flexibility of such a structure depends on how human annotations are gathered, retained, and accessed. To simplify such operations XR paradigms, such as VR interfaces for DT annotations and AR for physical ones (i.e., annotating objects and actions performed in the physical world), are employed. Such a framework supports the cross-accessibility of such information. In other words, those who work on the physical object may benefit from feedback and advice received from those who have worked with the DT and vice versa. An additional feature consists in visualizing a particular step of the process history, using vocal or visual markers that act as a revive triggers.

An assessment of the HCLINT-DT framework could be performed considering any scenario involving physical objects and their DTs where annotations are cross-provided, stored, and accessed. Such type of scenario is very general, as it spans from industrial to commercial to private settings. The assessment we performed considers a specific use case: memory preservation through annotations for family album photographs. Previous works have experimented with a HoloLens 2-based interface and Deep Learning (DL) paradigms for easy annotation and cataloging of pictures to revive the phenomena of exploring families' past[19,20]. Such works amount to a first step of the HCLINT framework: the annotation one. We expanded such works by developing a DT formed by the album's cyber-physical counterpart, along with the annotations made by the subject and the subject itself, all in the virtual realm. It is interesting to note that this use case shares traits common to other settings. For example, in manufacturing, workers often learn how to carry out an unknown process either by studying or by accessing the experiences shared by others[21]. The opportunity of anchoring the annotations to a physical object or its twin can foster the latter. This is not the only positive aspect though. Factors that influence employee involvement are, according to Marin et al.[22]: empowerment (sharing power with employees and increasing their level of autonomy), training (developing workforce shared abilities and a better understanding of the processes in which they participate), communication and remuneration. To assess whether a framework like HCLINT-DT could contribute to any of such factors in a productive setting, we also performed an on-sight observational analysis of an small and medium-sized enterprise in the business of building electrical systems.

This manuscript is organized as follows: in Section 2 we revise the works related to the framework introduced in Section 3. A possible instance is provided in Section 4, where we provided a practical use case, validated in Section 5. Section 6 presents a possible adaptation of the framework to a specific industrial case through an on-the-field observational study. Finally, Section 7 anticipates some future work on the VR interface, and Section 8 provides insights into the utility and value of this framework, along with possible future work directions.

2 Related works

DTs are increasingly adopted in research and industry[2,3,6–8], aiming at replicating, twinning, or mirroring some physical entity. Interestingly, in this context, a central role is progressively being earned by eXtended Reality (XR) paradigms, the umbrella term that groups together Virtual, Augmented, and Mixed Reality (VR, AR, and MR). Thanks to XR it is possible to manipulate DTs, directly influencing the physical world and vice versa[6–8]. A guide in such space can be found in [23], where the authors focus on reviewing AR/MR remote collaboration contributions to physical tasks. Among the future research issues they sketch for this domain, they mention multi-modal interaction, hybrid interfaces, and industrial AR-based remote collaboration, also citing DTs. In the following, we will focus on the research that falls closest to ours, using XR paradigms to maintain and sustain the growth and sharing of collaborative intelligence.

In [24], the authors proposed VirCA, namely the Virtual Collaboration Arena, to implement a complex vision by adopting a shareable and fully customizable 3D virtual workspace as a central idea. They aim to enable users, not necessarily co-located, to collaboratively create ideas and then design and implement them in a shared virtual space.

The authors of [25], instead, focused on providing a surgical telementoring tool where annotations are superimposed directly onto the surgical field using an AR simulated transparent display. With their system, annotations stick to the surgical field as the trainee display moves, and the surgical field deforms or becomes occluded.

Finally, in [26] an on-spot technician-manufacturer remote maintenance tool is proposed. The system can

record the malfunctions reported by end-users and provide, using an AR display, the maintenance actions suggested by an expert.

Concerning these contributions, we here concentrated on how humans could help others, adopting a Human Collaborative Intelligence (HCLINT) approach that assigns a central role to DTs. An HCLINT could help humans in supporting other humans in their activities. As reported by Chen et al.[18], HCLINT *involves an extensive collaboration of different team members to solve problems while giving a non-stop real-time learning opportunity*. Following this line of thought, we also resorted to a well-known knowledge transfer strategy commonly adopted by humans: asynchronous, persistent annotations. However, we moved a further step in this process: in our contribution, we support providing and sharing human annotations made of text, voice, or videos aligning both the physical and the virtual worlds utilizing XR (AR + VR in our case) paradigms. This amounts to a unique way to make humans learn from each other, fostering efficiency. Indeed, providing seamless exploitation of annotations provided via AR and VR paradigms can empower those who interact with the physical object, receiving feedback from those who exploit its twin and vice versa.

Hence, when compared to the works cited in this Section, ours is the only one that aims at designing an HCLINT module to empower DTs through annotation while maintaining contact with both the physical and virtual space by means of XR paradigms. In addition, we here pursued a user-centric approach by providing an assessment adopting two different strategies, one using Technology Acceptance Model (TAM) constructs and a second conducting a short-term observational study within a manufacturing plant[27,28]. In the following, we describe how we propose supporting such a collaborative environment through XR paradigms.

3 HCLINT-DT framework: exploiting human collaborative intelligence to empower Digital Twins

In this section, we describe the framework that aims at exploiting Human Collaborative Intelligence (HCLINT) to empower the five-dimensional Digital Twin (DT) model[4], as depicted in the here introduced Figure 1. For the sake of clarity, in Table 1 are reported the principal acronyms, along with the corresponding full-name, used in Sections 3 and 4.

Digital Twin model. The five-dimensional DT model adopted in this work builds upon the framework

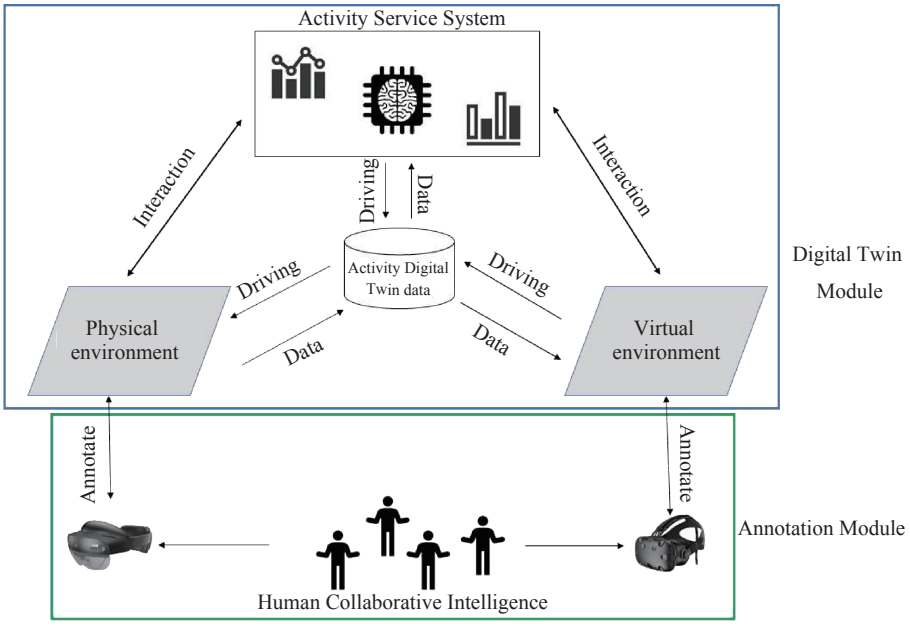

Figure 1 Main components of the HCLINT-DT framework.

Table 1 Table of acronyms along with their full textual description

Acronym	Full-name	Acronym	Full-name
ADTD	Activity Digital Twin Data	HCLINT	Human Collaborative Intelligence
ASSYST	Activity Service System	AM	Annotation Module
PE	Physical Environment	XR	eXtended Reality
VE	Virtual Environment	AR	Augmented Reality
DT	Digital Twin	VR	Virtual Reality

presented in[4], comprising the following components: (a) a Physical Environment (PE) that includes a series of entities, such as humans, machines, and documents; (b) a Virtual Environment (VE) consisting of models built in multiple dimensions, including geometry, physics, behavior, and rules, evolving according to the PE, and the (c) Activity Service System (ASSYST). ASSYST amounts to an integrated service module, which encapsulates the functions of data analytics, models, algorithms, etc., into sub-services, and combines them to form composite services for specific demands from the PE and the VE. Finally, (d) the Activity Digital Twin Data (ADTD) includes the PE, VE, and ASSYST data, their aggregations, and the existing modeling methods, optimizing and empowering both the PE and VE. The data in ADTD communicates in real-time with all other modules to eliminate possible islands of information to provide a comprehensive, synchronized, and consistent vision.

Annotation Module. The *Annotation Module* (AM) emerges as a companion to the five-dimensional DT module. The AM acts as a plug-and-play module that exploits HCLINT together with XR paradigms. In practice, each user that interacts with the PE or the VE can read or produce annotations. Such annotations can originate from both AR and VR. In the first case, a user wearing an AR head-mounted display (e.g., the Hololens) can: (a) individuate those real-world objects that also possess a cyber-physical counterpart in the VE, and (b) provide annotations, with the production of multimedia content (e.g., textual, visual and vocal data). The ADTD and the ASSYST modules process and replicate such annotations, making them available to the corresponding elements in the VE. The VR setting offers instead the possibility to directly provide annotations in the VE, exploiting a VR head-mounted display. The updates shared in VR are processed by the ADTD and the ASSYST modules to export such annotations to AR. The proposed AM is agnostic concerning the specific type of DT (object, human, or process).

To visually depict the operation mechanism of the AM, and how it interacts with the reference DT model, we introduce the novel Figure 2. As illustrated the AM works in three stages, before, during and after inserting an annotation (or not). In this Figure, the blue, purple, yellow, and green blocks represent PE, VE, ASSYST, and AR/VR interfaces, respectively. Their operations and interactions are supported by the ADTD.

Before reading or producing an annotation, the system recognizes the objects a user watches. To this end, a snapshot of the user view is taken from the PE or the VE, depending on the interface the user decided to use (AR vs. VR). In the case of AR, the snapshot is a simple photo or video taken from the camera, with VR it includes the set of objects that lie in the viewing frustum. At this point, an object recognition service (provided by the ASSYST module) is invoked. In the case of AR, such a module employs one or more Deep Learning-based object detectors to identify known objects, along with their spatial coordinates. For VR, instead, the recognition consists in examining meta-data associated with the set of considered objects. Once the objects in the user view are recognized, the ASSYST module retrieves the related annotations, if any exist. Meanwhile, the objects recognized in the user's view are highlighted to suggest which are active for possible interactions. Finally, when the user interacts with one of those, a menu overlays on the interface, providing the list of existing annotations.

Now, a given user could decide to provide a new annotation or visualize one of the existing ones (the annotations could be many). In the first case, the user exploits virtual keyboards or recorders based on the type of annotation (textual vs. vocal). In the second, the ASSYST module stores the new annotation and the identity

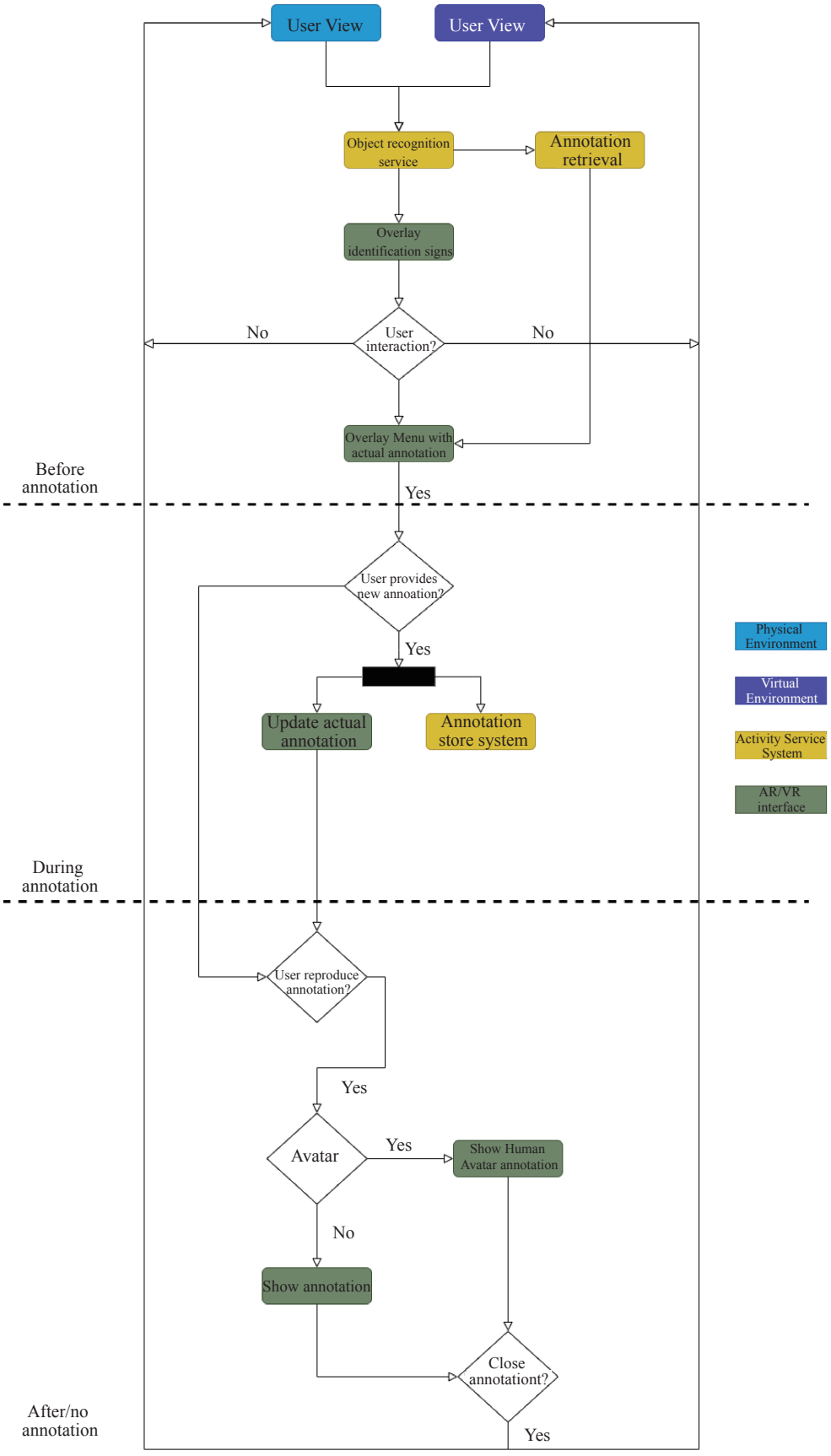

Figure 2 HCLINT-DT mechanism workflow: interactions between AM and the DT model.

of its author. Finally, further annotations could be either provided or reproduced.

When the user decides to reproduce an actual annotation, it may simply visualize a textual item or pick a more natural way of acquiring the information inside. To this aim, a user can listen to the contents of a given annotation directly from the avatar of the person that shared it. After this step, the process loops from the start.

4 Experimenting with Digital Twinning: a HCLINT-DT use case for family photo albums

In this section, we provide a practical use-case for the HCLINT-DT module: the DT of a family photo album annotation process. Family photo albums provide an unrepeatable chance to revive old memories about social events, affections, relatives, friends, special events, etc.[29]. Throughout the 20th century, people printed photos and collected them in family albums. Despite the spread of digital photography and social media, people still look back and discover their families′ pasts, often sharing think-aloud thoughts and memories that typically survive for the time of the conversations where they were exchanged[19,20,30]. Digital technologies may provide viable ways of retaining memories, annotating, and reviving such elements in an easy and meaningful way.

Applying the HCLINT-DT framework it is possible to experiment its ability of storing and providing relevant annotations. Often, when a photo is placed in an album, a few annotations are written on its back. Those who will browse those pictures afterward will be able to discover what that picture portrayed thanks to those annotations. This amounts to a typical example of HCLINT, where the production and consumption of information occur collaboratively.

We hence designed the DT of the family photo album, applying the HCLINT-DT module and implementing a two-sided AR and VR application for annotation sharing.

4.1 AR interface

The AR interface resorts to models developed in previous research[19,20], where a system was developed for the digitization and cataloging of collections of family album photographs exploiting: the HoloLens 2[31] as a wearable device, AR paradigms to implement the interface, and Deep Learning models to catalog family album photos. In particular, we fine-tuned a well-known object detector, namely YOLOv5[32], to identify the pictures within a given user′s view. Such pictures are then classified according to socio-historical labels, using the IMAGO models provided in[33], namely IMAGO-DATING and IMAGO SOCIO-HISTORICAL. These models provide the prediction of the date and the socio-historical context of an analog family album photo, respectively.

We decided to extend the AR system introduced in such work, taking YOLOv5 as a picture detector and adding a module to identify which pictures are already in the ADTD[34]. In this way, it is possible to recognize the pictures that are included in the database of the DT of the photo album and retrieve the associated annotations. Hence, all models run as part of the ASSYST module defined in Section 3.

The AR interface comprises three modules: picture detection and matching, annotations retrieval, and interaction menu. Firstly, the AR interface captures the user view to individuate any family album photos. At this point, the pre-trained YOLOv5 object detector predicts the bounding box coordinates of the different pictures in the scene and crops them accordingly. The procedure is graphically reported in Figure 3.

The cropped pictures are then passed to an image-matching algorithm to check whether they carry any annotations. If this is the case, their annotations are retrieved. To this aim, we employed the SIFT[35] algorithm for feature matching. Otherwise, the user could be the first to annotate them. The workflow appears in Figure 4.

In any case, once the recognition and retrieval steps are completed, the user can revive the annotations made by previous users on the considered photos (if present).

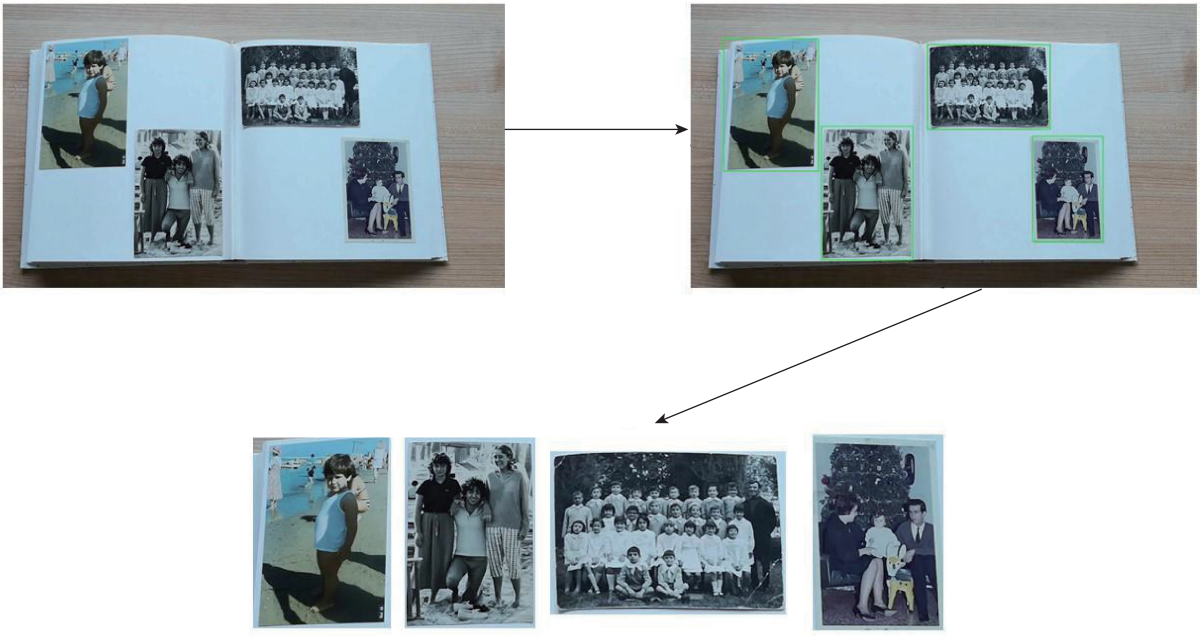

Figure 3 Execution of pre-trained YOLOv5 to individuate and crop pictures.

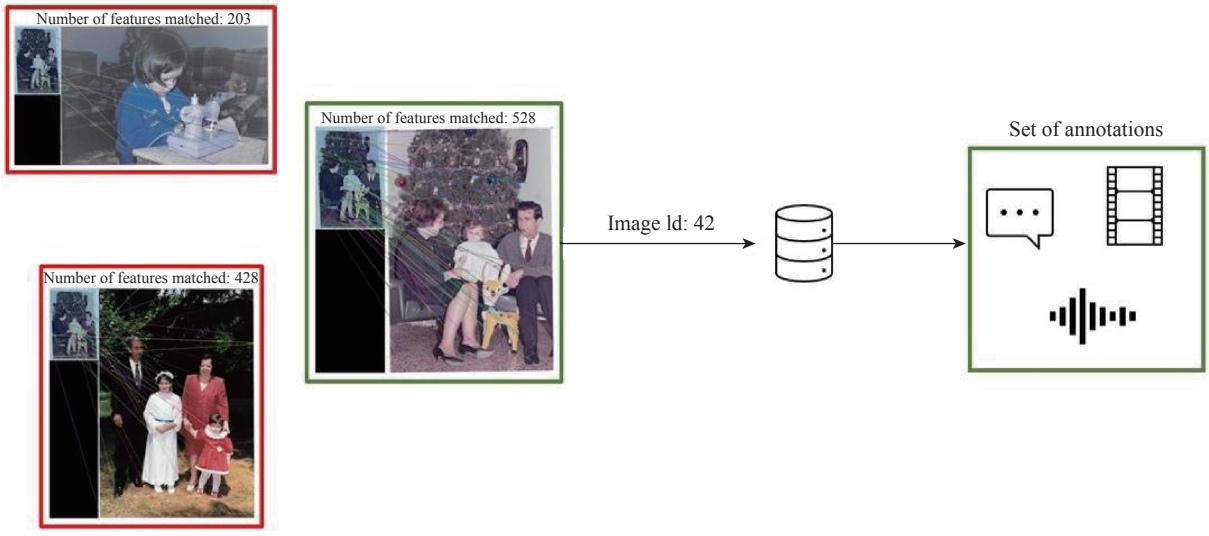

Figure 4 Example of SIFT execution, ranking and subsequent database search for annotation retrieval.

The AR menu was thought to be as simple as possible to let the user focus on pictures, without being distracted. Such menu is reported in Figure 5.

Recalling that, in the previous step, the bounding boxes of pictures were obtained, it is possible to create an invisible interaction area, amounting to the rectangles enclosing the photos. Exploiting this method, when the user touches one of the pictures, the menu appears. At this point, the user could decide to write a new annotation or reproduce an existing one. In Figure 6 the procedure to create a new annotation is reported, considering the textual type. In practice, the user interacts with the menu by touching the New Annotation button and then decides if s/he wants to leave a textual, vocal, or video annotation. For each type of annotation, a different menu appears. Considering the textual one, a virtual keyboard, used as an input device, will materialize next to the menu. Considering instead vocal and video inputs, the user could register a vocal note about a particular picture, associating its 3D avatar. The latter may be implemented by adopting Deep Learning

Figure 5 AR interface main Menu.

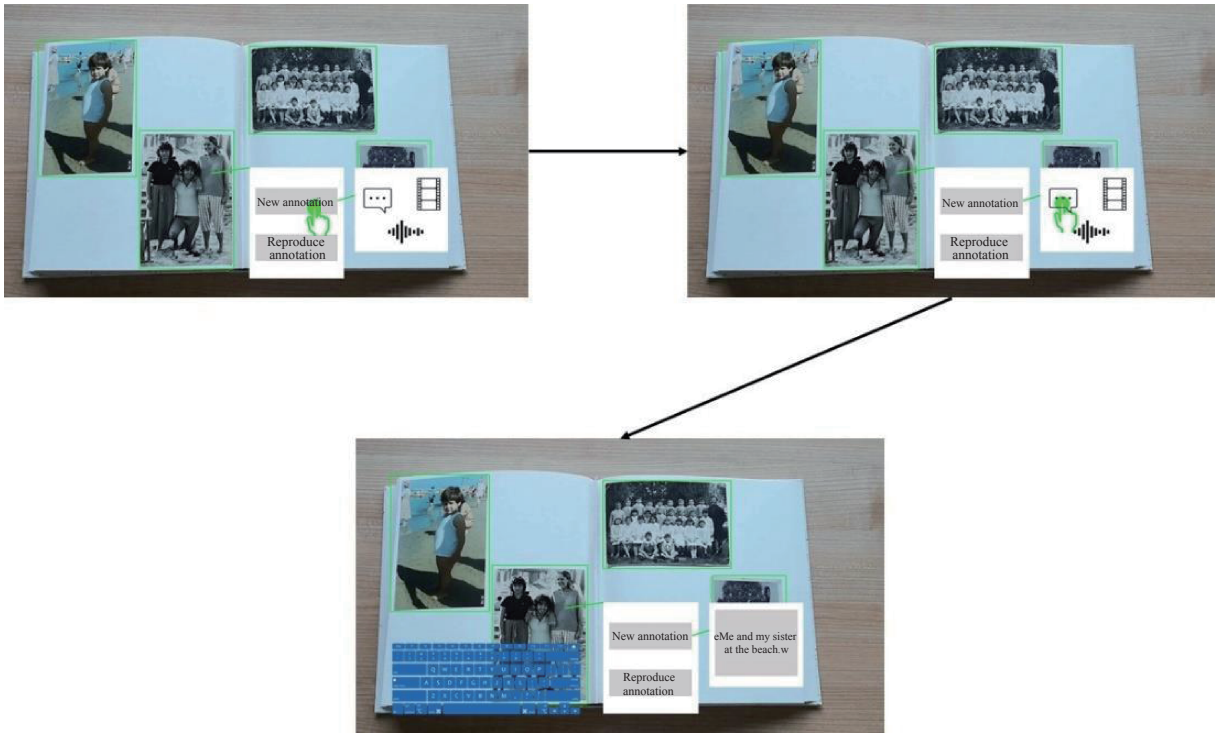

Figure 6 AR interface Menu: annotate a picture.

models like the one provided in [36] and tools like [37] (included in the ASSYST).

All *new annotation*s are immediately sent to the ADTD module to support their access to the VE. After introducing any annotations, the user could also reproduce an existing one. In this case, as reported in Figure 7, s/he may select the Reproduce Annotation, and then the kind of annotations s/he wants to reproduce, finally picking one from the list. The Figure exhibits the textual annotation case.

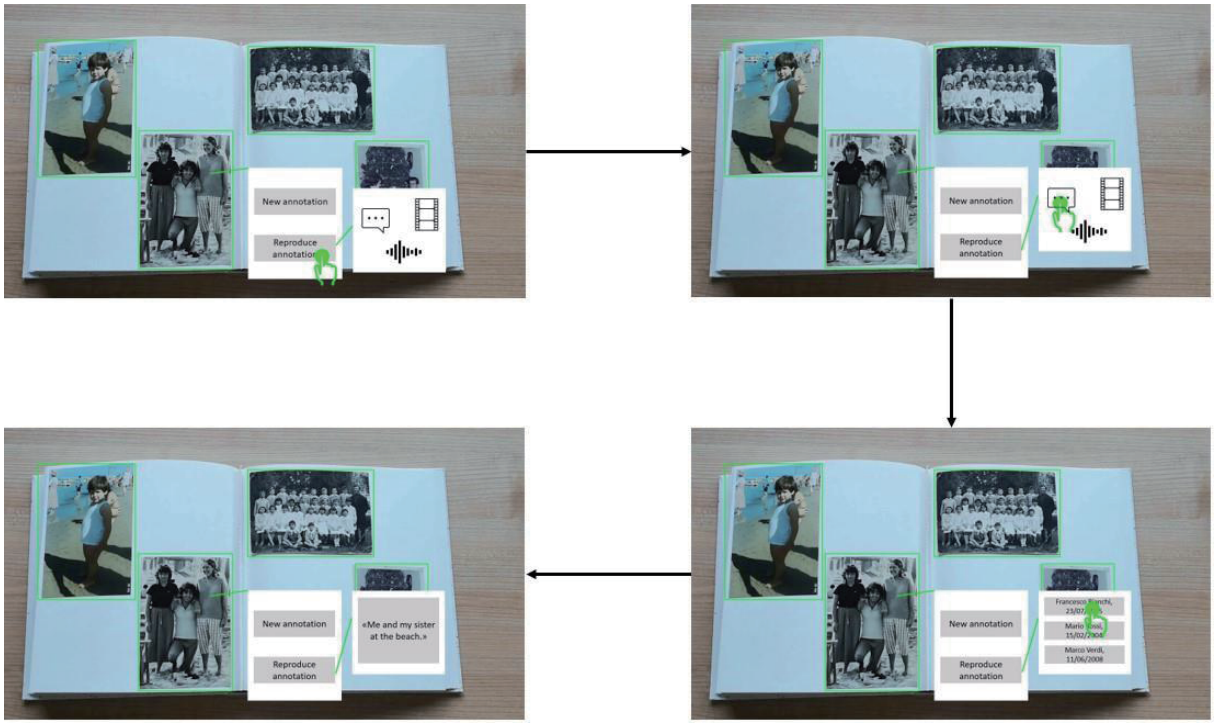

Figure 7 AR interface main Menu: reproduce textual annotations of a picture.

4.2 VR interface

The VR application carries out the same tasks carried out by the AR interface but exploits VR paradigms. All the environment, objects, and interactions were developed using Unity[38]. In Figure 8 we report the initial user view when s/he just entered the environment.

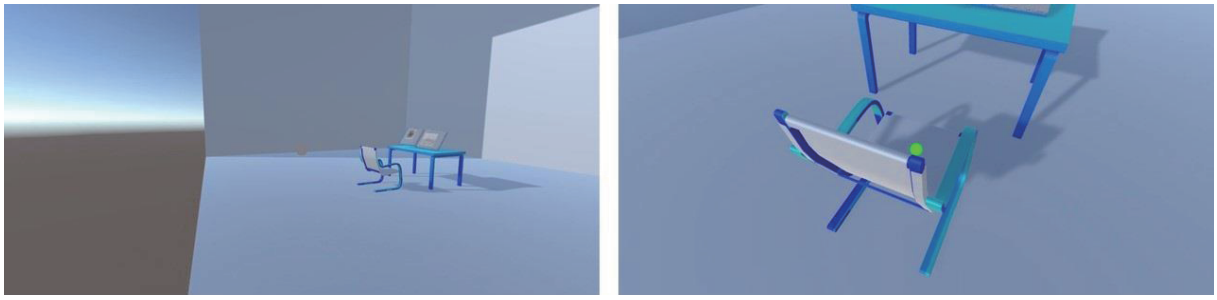

Figure 8 Initial VR user view and user seat.

As depicted, the environment is straightforward: there is an empty room with a table, a chair, and a family photo album showing two pictures. The visualized pictures could be changed with a shift-like command as if the user were browsing a family album. All user interactions use ray casting, allowing the selection of items and actions by a simple point-and-click mechanism. This design choice is based on the Google VR SDK and was adopted to adapt such an environment to multiple mobile VR devices (e.g., smartphones)[39].

The main activity that the user could carry out is browsing the family album to revive memories behind the picture. Also, the pictures are interactive: if the user clicks on one of them, the main menu appears. As for the AR interface, the menu provides the possibility to create a new annotation or reproduce already existing ones about that particular photo (as graphically reported in Figure 9).

As previously stated, the annotation could be textual or vocal/visual. Starting from the first, once the Textual

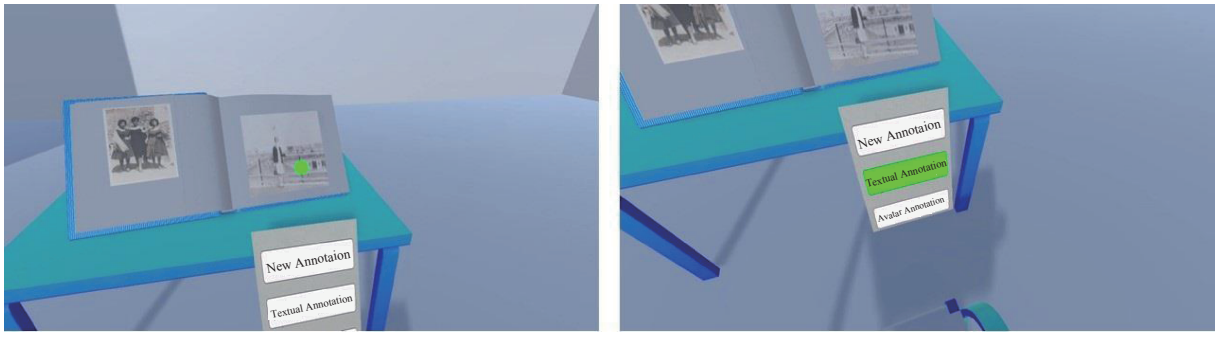

Figure 9 VR family album and Menu.

Annotation button is selected, the menu shows all the textual annotations made by other users. In case there is a unique annotation, this is directly reported on top of the picture, as depicted in Figure 10.

Figure 10 Textual annotation of a particular picture, reported in its original language.

Considering now vocal and visual annotations, we here report the results of their analysis. Indeed, as explained in Section 4.1, the user, after recording a vocal note, could also produce its 3D-avatar. This process results in a holistic Avatar Annotation, that can be reproduced in the virtual realm, as reported in Figure 11.

In addition, the user could also decide to provide a new annotation for the considered picture. As for the AR interface, s/he could provide a textual and a vocal/visual annotation following the same steps but exploiting VR paradigms. For example, considering the vocal annotation process, the user should select the New Annotation button and then the Vocal mode. A Sub-Menu will appear as reported in Figure 12. At this point, s/he could click Start recording while contextually talking. Once finished, the Stop recording button provides the possibility to save the vocal note. All the produced annotations are then sent to the ADTD module, which

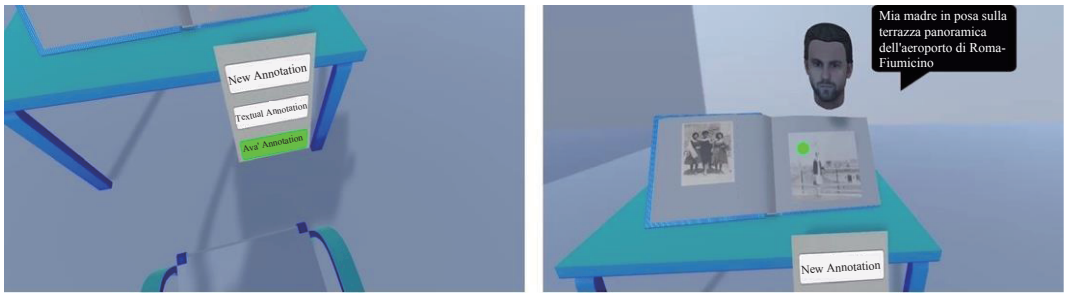

Figure 11 Reproducing avatar annotation in VR.

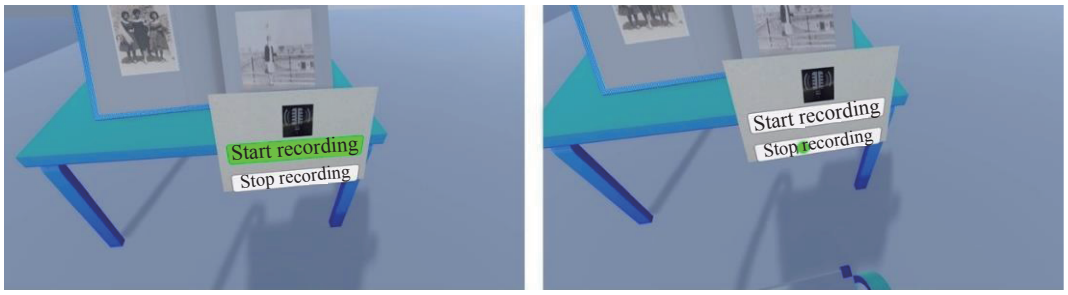

Figure 12 Vocal Sub-Menu and recording mechanism.

will relay them to the ASSYST module and to the PE (i.e., and so the AR interface).

5 Assessment model and results

We report the results obtained with the assessments of the AR interface as a tool employed to insert annotations. The outcomes presented here represent the experience of interacting with the physical object (i.e., a family photo album) via an AR interface. One of the primary use cases for our framework would be the one involving a user making annotations on real objects. Others should be able to easily access that information, even if not in the physical place, from the VE. The evaluation was performed with an online survey in which the participants were first asked to watch a video of how the AR interface worked. The survey was administered to a group of 30 participants. The group included a gender-balanced number of participants: 15 males and 15 females. The average age of the participants corresponds to 28.17 and a standard deviation of 3.12 (youngest and oldest 22 and 37 years old, respectively). The number of participants has been chosen as a trade-off between the necessity of acquiring sufficient feedback data from a population and the time spent for the evaluation phase[40,41].

Assessment model. The *assessment model* was designed to evaluate: (a) the ease and intention of use of such an interface, and (b) its usefulness/adoption. To design such a survey, we took inspiration from the Technology Acceptance Model (TAM)[42] used to measure user intentions in terms of their attitudes, subjective norms, perceived ease of use, and usefulness. From the TAM, we derived the following questions:

I1. I found the new interface easy to understand (5-point Likert scale);

I2. I would prefer watching an Augmented Family Photo Album instead of a normal one (5-point Likert scale);

I3. I appreciated the automatic identification of the pictures (5-point Likert scale);

I4. Would you use this AR interface to share your annotations? (yes/no question);

I5. I enjoyed the overall experience (5-point Likert scale).

The I1 sentence aims at evaluating the easiness of use of the AR interface design; item I2 lets us understand whether the users prefer to live an augmented experience or a classical one. Item I3 aims at measuring the usefulness of one of the main features of the AR interface: the automatic identification of pictures utilizing YOLOv5 and SIFT. Then, I4 serves the purpose of understanding how much our users want to use this AR interface to share their memories, and so the annotations of their pictures. Finally, through I5, we ask for a broad evaluation. As reported, four out of five items are measured with a five-point Likert scale, except for item I4. This item was formulated as a Yes/No question because we wanted to emphasize the direct intention of the users to use the annotation system.

Results. In Table 2 are reported the means along with the standard deviation obtained by surveying our subjects on the proposed items. The internal consistency of the questionnaire was verified by adopting the

Table 2 Mean and Standard deviation for all the considered items

Item	I1	I2	I3	I4	I5
Mean	4.37	3.70	4.37	0.77	4.30
Std	1.00	1.06	0.93	0.43	0.88

Cronbach's alpha index, which corresponds to the Kuder-Richardson Formula 20 (KR-20) in the case of binary choice questions, such as our Q-items. The returned Cronbach's alpha index corresponds to 0.751, which can be considered reliable ($\geqslant$ 70, as indicated by [43]). This is a valuable outcome, considering that some items could be divisive (e.g., I2 vs. I3). After checking the internal consistency and validity of our questionnaire, we report in Table 2 all the means and standard deviations obtained by surveying our subjects on the proposed items along with the item response histograms in Figures 13 and 14.

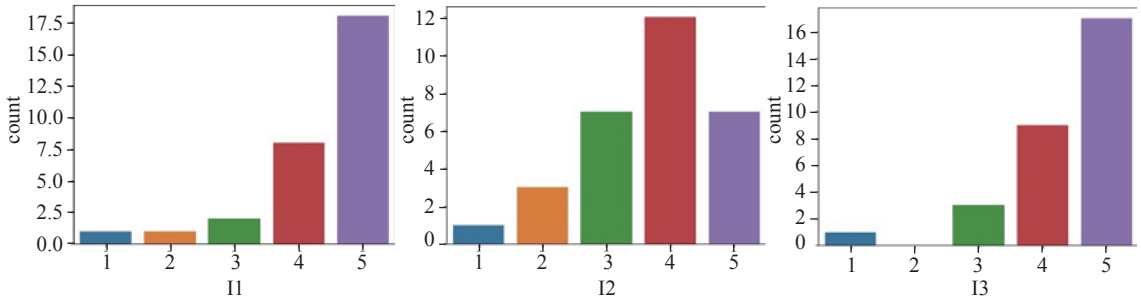

Figure 13 Histograms for answers in 5-point Likert scale items I1, I2 and I3.

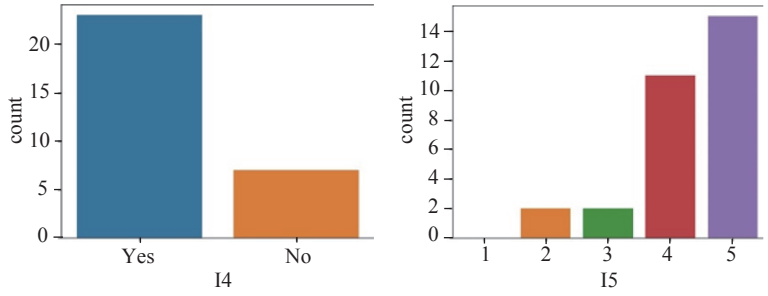

Figure 14 Histograms for answers Yes/No and 5-point Likert scale items I4 and I5.

From Table 2 and Figure 13, it is evident that there is a general agreement about the ease of use of the interface design (I1), but also a not so positive outcome when compared to the use of modern technologies in the given application scenario (I3). However, this last outcome contrasts with answers for item I2, where we can only appreciate a partial agreement in the preference of watching a family photo album through the lens of modern technology. This is probably because some of the respondents continue to prefer reviving their old memories physically without filters. For what concerns instead of the usage of the proposed AR interface while annotating their picture (I4), Table 2 and Figure 14 exhibit an agreement level that supports the initial aim of our application. Finally, scores from I5 highlight that all the subjects appreciated the AR interface.

To further confirm our results, and test the statistical significance of the obtained answers, we performed a one-sample *t*-test[44] over all the five-point Likert scale items. The one-sample t-test compares the mean to a hypothesized mean value as long as the samples follow a normal distribution. Since the sample size is thirty, it approximates the standard normal distribution according to the central limit theorem[45]. However, the classical two-tailed one-sample *t*-test does not highlight the direction of the difference between the sample mean and the hypothesized mean value. For this reason, we performed a one-tailed one-sample *t*-test in which the null

hypothesis H_0 assumes that the mean (μ) is lower or equal to the fixed mean value, while the alternate H_1 that (μ) is higher. The compared value was set as three for the five-point Likert scale items since any value greater than three demonstrates a partial agreement in a five-point Likert scale. We set the parameter $\alpha = 0.05$ as the significative threshold for the *p*-value. The results of the performed test on all the considered five-point Likert scale items are reported in Table 3.

Table 3 One-tailed one-sample t-test performed over all the considered 5-point Likert scale items. For each of them the null hypothesis can be rejected

Item	I1	I2	I3	I5
Value	7.49	3.63	8.06	8.12
P-value	1.48e-08	5.3e-04	3.4e-09	3e-09

From Table 3 it is evident that, for all the considered items, the null hypothesis H_0 can be rejected, considering both the significative threshold for the *p*-values, but also the critical *t*-values[46], thus confirming the analysis carried out so far.

For what concerns, instead, the unique Yes/No question item (I4), we performed a Binomial Test returning the probability for the assumption that the observed frequencies are equal to the expected frequencies[47]. Also, in this case, we adopted a one-tailed perspective fixing the expected probability of the positive outcome to 0.6 (positive agreement). The null hypothesis H_0 assumes so that the probability $P(\text{I4} = 1) \leq 0.6$, while the alternate hypothesis H_1 assumes that $P(\text{I4} = 1) > 0.6$. We set the parameter $\alpha = 0.05$ as the significative threshold for the *p*-value. The returned proportion estimate value (0.77), and the *p*-value = 0.044, demonstrate the statistical significance of our test and so the positive outcome of the Item I4.

Finally, it is worth noticing that all the questions provided to our subjects were very general: this choice was taken to evaluate an abstract scenario, that can be applied in any context in which a person is watching an object and wants to take/retrieve annotations on it. This can be seen as a simplified form of any watch-and-annotate scenario, where the user can share but also retrieve annotations from other remote users, taking full advantage of HCLINT.

6 HCLINT-DT: a short term observational analysis in an industrial setting

As anticipated, the HCLINT-DT may be generally adapted to different contexts. To explore this possibility, we carried out a short-term observational study at Elettrotecnica Imolese S.U.R.L.[48], whose mission is to produce industrial electrical systems, starting from their design to the final installation and testing. In particular, we observed and interviewed professional workers in their everyday activities to understand how HCLINT-DT could improve their processes from a lean manufacturing perspective[22,49]. We concentrated on how to improve *Empowerment*, *Communication*, and *Training* factors using HCLINT-DT where the main worker collaborations activities take place. Observing professional workers during their activities, we eyed that human collaborations take place while assembling, installing, and testing electrical switchboards. In particular, while assembling electrical switchboards, professional workers rely on portable computers, which allow them to visualize documents describing electrical schemes and circuits. Workers can make annotations about missing materials, or mistakes or scribble checkmarks on components that they correctly assembled. All these annotations are shared, utilizing cloud technologies, to engineers that will provide corrections to the electrical schemes, again annotating the same documents. These activities produce a loop that ends when the assembling phase ends. This first processes are visually represented in Figure 15. After the installation, the testing step takes place when the professional testers verify, following strict protocols, that the switchboards are correctly installed, analyzing each of the components and their interconnections. For each component, information is annotated on the same shared document with positive or negative checks. Regardless of the outcome (success

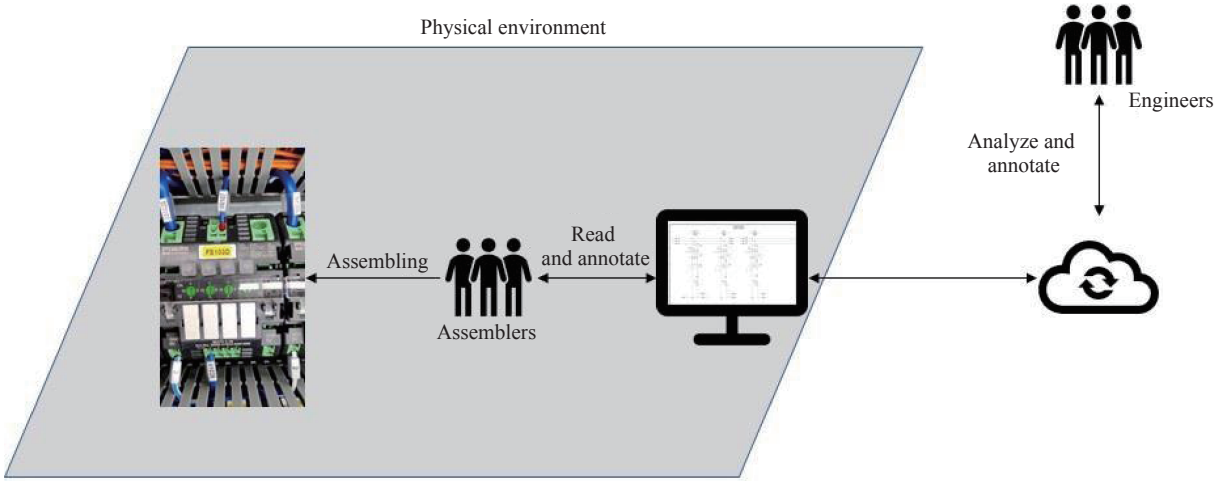

Figure 15 In the assembling phase, assemblers read and annotate shared documents of electrical switchboards and engineers eventually analyze and correct errors or suggestions made by the assemblers.

or failure), the annotated documents are shared back with the assemblers.

Considering this context of use, the HCLINT-DT framework, and the three factors that most influence lean manufacturing, we designed a lean manufacturing optimization process, taking inspiration from the model provided by Gupta et al.[49]. Firstly, we identified the weaknesses of the existing process. We found out that the main one amounts to the superfluous time spent by the assemblers while: (a) physically moving from the portable pc to the physical switchboards, and (b) searching for the correct page in the document to read the instruction or annotate particular components and its connections. We then sketched the implementation of the HCLINT-DT to erase these time-wasters by simultaneously improving *Empowerment*, *Training* and *Communication* factors. In practice, we started defining the DT of an electrical switchboard by considering all its components and interconnections. As for the considered family album photos use cases, an AR interface (described in Section 4.1) would recognize each component in the real world, providing the possibility to make or read annotations by different users. In this case, however, the read/write annotation process would be mediated by digital documents (e.g., pdfs). This means that, when a component is recognized, the document pages containing relevant information will be visualized in the user′s view and manipulated by employing AR paradigms. In the virtual realm, instead, an exact reproduction of the switchboards along with all the annotations posed by different figures is defined. This approach should improve on the existing process. Firstly, the AR interface would run on an AR device like the Hololens, so the assemblers could keep going with their work without moving from the electrical board to the personal computer. Secondly, object recognition would prevent wasting time searching for the instruction/annotations regarding a particular component by immediately visualizing all the info in the shared document. The adaptation of the HCLINT-DT framework for this use case is visually reported in Figure 16.

In addition, thanks to the adoption of the HCLINT-DT framework, the workers′ *Empowerment* is provided with the use of state of the art tools and XR devices (e.g., HoloLens 2 and Htc-Vive), as confirmed by the assemblers, engineers, and testers which were interviewed during our analysis. The *Communication* improvement is given by the natural working mechanism of HCLINT-DT shared annotations. Finally, *Training* is a natural consequence of the same HCLINT-DT shared annotation mechanism since workers could analyze previously annotated documents to learn from others' experiences.

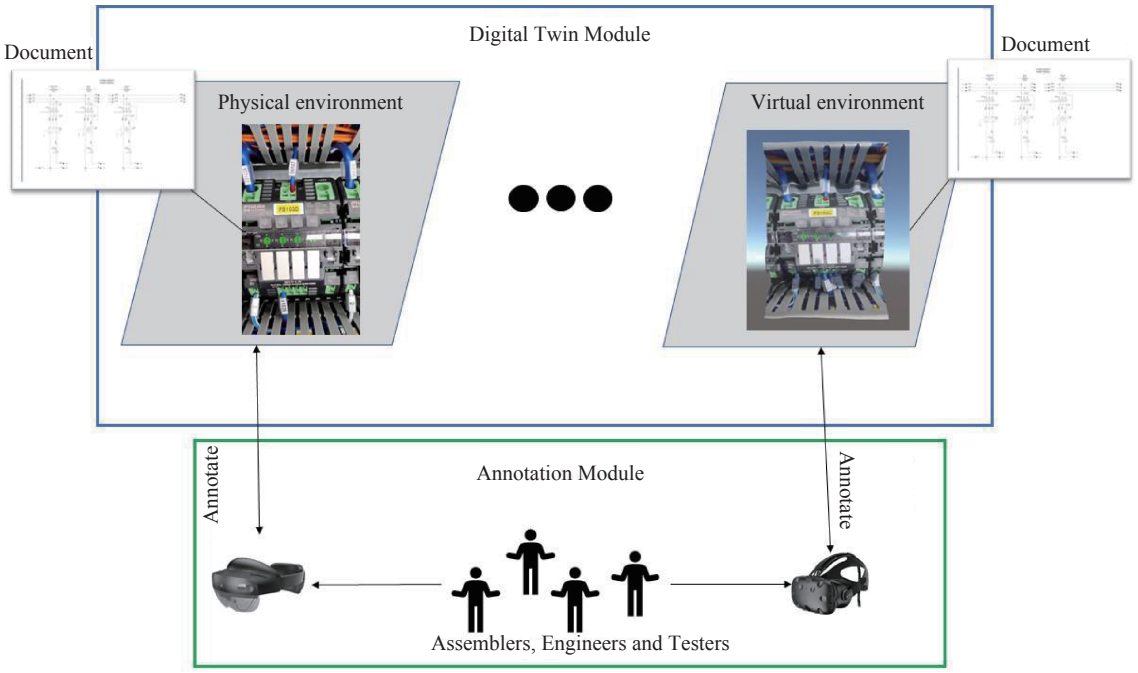

Figure 16 HCLINT-DT framework adapted to the Elettrotecnica Imolese use case: physical components of electrical switchboards are recognized giving the chance to read or write annotation that are mirrored in its DT in the virtual space, that also provides the same possibilities.

7 Future works

In Section 4.2 we introduced a first version of the Virtual Environment hosting a Digital Twin (DT) of a family photo album empowered by human annotations. The main functionalities of such an interface regard the fruition of human annotations from anywhere, at any time. However, the annotation retrieval methods introduced use the classic "point-and-click" mechanism. This retrieval method could not be sufficient to satisfy the needs of the user base that intend to explore human annotations. Further investigations about how to improve annotations retrieval are required. In particular, it is necessary to carry out a thorough assessment study of the VE. In addition, the possible impact of natural language processing technologies, such as speech-to-text and text parsing (to provide vocal commands), named entity recognition (to recognize the person a user is referring to), and word embeddings (to ease the query of a particular annotation) should also be evaluated. Finally, we will study how annotations may automatically influence the physical space, exceeding hence their traditional role.

Considering now the here introduced HCLINT-DT framework (Section 3), possible future directions of work include the integration of crowd intelligence technologies, one of the most promising in the field of AI and DTs[50,51]. Crowd intelligence emerges from the collective intelligent efforts of massive numbers of autonomous individuals, who are motivated to carry out challenging computational tasks under an Internet-based organizational structure[50]. Many Internet applications, such as Wikipedia, Web Q&A, and sharing economy, have been exploited into pools of talented crowds demonstrating significant progress beyond the traditional paradigms. It is now interesting to consider that crowd intelligence systems interweave crowd and machine capabilities to address challenging computational problems[50]. Indeed, the outcomes of crowd intelligent tasks, including data collection and annotation, could help train AI algorithms and models. Matter of further investigations will also include understanding how the HCLINT-DT framework could support machines (learning) and the consequences of such processes on human activities.

All these further directions of work will be supported by the involvement of larger populations of users, to

provide more general assessments.

8 Conclusions

We introduced the HCLINT-DT framework to support the spread of human collaborative intelligence by leveraging DT and XR paradigms. To validate such an approach, we assessed a use case involving family photo albums through an online survey. The results show a general agreement on the ease of use of the AR interface and the overall experience, even if there was only a partial agreement in preferring AR. We also explored the adaptability of the proposed approach considering a use case drawn from a local industrial electrical engineering context. Here, the HCLINT-DT showed a good adaptability level. Further investigations could involve: (a) improving annotations retrieval, (b) analyzing how this annotation system could support machines (learning), and, (c) understanding more deeply the impact of such systems on human activities. Finally, instances of this framework could be created for additional areas and domains, other than the considered ones (e.g., education, marketing).

Declaration of competing interest

We declare that we have no conflict of interest.

References

1 Barricelli B R, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access, 2019, 7: 167653–167671
DOI: 10.1109/access.2019.2953499

2 Tao F, Zhang H, Liu A, Nee A Y C. Digital twin in industry: state-of-the-art. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2405–2415
DOI: 10.1109/tii.2018.2873186

3 Elayan H, Aloqaily M, Guizani M. Digital twin for intelligent context-aware IoT healthcare systems. IEEE Internet of Things Journal, 2021, 8(23): 16749–16757
DOI: 10.1109/jiot.2021.3051158

4 Tao F, Zhang M. Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. IEEE Access, 2017, 5: 20418–20427
DOI: 10.1109/access.2017.2756069

5 Muñoz-Saavedra L, Miró-Amarante L, Domínguez-Morales M. Augmented and virtual reality evolution and future tendency. Applied Sciences, 2020, 10(1): 322
DOI: 10.3390/app10010322

6 Mukhopadhyay A, Reddy G S R, Saluja K P S, Ghosh S, Peña-Rios A, Gopal G, Biswas P. Virtual-reality-based digital twin of office spaces with social distance measurement feature. Virtual Reality & Intelligent Hardware, 2022, 4(1): 55–75
DOI: 10.1016/j.vrih.2022.01.004

7 Rasheed A, San O, Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. IEEE Access, 2020, 821980–22012
DOI: 10.1109/access.2020.2970143

8 Sepasgozar S M E. Digital twin and web-based virtual gaming technologies for online education: a case of construction management and engineering. Applied Sciences, 2020, 10(13): 4678
DOI: 10.3390/app10134678

9 Schroeder G, Steinmetz C, Pereira C E, Muller I, Garcia N, Espindola D, Rodrigues R. Visualising the digital twin using web services and augmented reality. In: 2016 IEEE 14th International Conference on Industrial Informatics. Poitiers, France, IEEE, 2016, 522–527
DOI:10.1109/indin.2016.7819217

10 Zhu Z, Liu C, Xu X. Visualisation of the Digital Twin data in manufacturing by using Augmented Reality. Procedia CIRP, 2019, 81: 898–903
DOI: 10.1016/j.procir.2019.03.223

11 Qiu C, Zhou S, Liu Z, Gao Q, Tan J. Digital assembly technology based on augmented reality and digital twins: a review. Virtual Reality & Intelligent Hardware, 2019, 1(6): 597–610
DOI: 10.1016/j.vrih.2019.10.002

12 Josifovska K, Yigitbas E, Engels G. A digital twin-based multi-modal UI adaptation framework for assistance systems in industry 4.0. In: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, 398–409
DOI: 10.1007/978-3-030-22636-7_30
13 Ma X, Tao F, Zhang M, Wang T, Zuo Y. Digital twin enhanced human-machine interaction in product lifecycle. Procedia CIRP, 2019, 83: 789–793
DOI: 10.1016/j.procir.2019.04.330
14 Goldberg K. Robots and the return to collaborative intelligence. Nature Machine Intelligence, 2019, 1(1): 2–4
DOI: 10.1038/s42256-018-0008-x
15 Hackman J. Collaborative intelligence: using teams to solve hard problems. 2011
16 Epstein S L. Wanted: collaborative intelligence. Artificial Intelligence, 2015, 221: 36–45
DOI: 10.1016/j.artint.2014.12.006
17 Wilson H J, Daugherty P R. Collaborative intelligence: humans and ai are joining forces. Harvard Business Review, 2018, 96, 114–123
18 Chen Y, Lee G M, Shu L, Crespi N. Industrial Internet of Things-based collaborative sensing intelligence: framework and research challenges. Sensors, 2016, 16(2): 215
DOI: 10.3390/s16020215
19 Stacchio L, Hajahmadi S, Marfia G. Preserving family album photos with the HoloLens 2. In: 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. Lisbon, Portugal, IEEE, 2021, 643–644
DOI: 10.1109/vrw52623.2021.00204
20 Stacchio L, Angeli A, Hajahmadi S, Marfia G. Revive family photo albums through a collaborative environment exploiting the HoloLens 2. In: 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct. Bari, Italy, IEEE, 2021, 378–383
DOI: 10.1109/ismar-adjunct54149.2021.00086
21 Gerber R. How do workers learn in their work? The Learning Organization, 1998, 5(4): 168–175
DOI: 10.1108/09696479810228469
22 Marin-Garcia J A, Bonavia T. Relationship between employee involvement and lean manufacturing and its effect on performance in a rigid continuous process industry. International Journal of Production Research, 2015, 53(11): 3260–3275
DOI: 10.1080/00207543.2014.975852
23 Wang P, Bai X, Billinghurst M, Zhang S, Zhang X, Wang S, He W, Yan Y, Ji H. AR/MR remote collaboration on physical tasks: a review. Robotics and Computer-Integrated Manufacturing, 2021, 72: 102071
DOI: 10.1016/j.rcim.2020.102071
24 Galambos P, Csapó Á, Zentay P, Fülöp I M, Haidegger T, Baranyi P, Rudas I J. Design, programming and orchestration of heterogeneous manufacturing systems through VR-powered remote collaboration. Robotics and Computer-Integrated Manufacturing, 2015, 33: 68–77
DOI: 10.1016/j.rcim.2014.08.012
25 Andersen D, Popescu V, Cabrera M E, Shanghavi A, Gomez G, Marley S, Mullis B, Wachs J. Virtual annotations of the surgical field through an augmented reality transparent display. The Visual Computer, 2016, 32(11): 1481–1498
DOI: 10.1007/s00371-015-1135-6
26 Mourtzis D, Zogopoulos V, Vlachou E. Augmented reality application to support remote maintenance as a service in the robotics industry. Procedia CIRP, 2017, 63: 46–51
DOI: 10.1016/j.procir.2017.03.154
27 Yuniarto D, Helmiawan M A, Firmansyah E. Technology acceptance in augmented reality. Jurnal Online Informatika, 2018, 3(1): 10
DOI: 10.15575/join.v3i1.158
28 Dünser A, Hornecker E. An observational study of children interacting with an augmented story book. In: Technologies for E-Learning and Digital Entertainment. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, 305–315
DOI: 10.1007/978-3-540-73011-8_31
29 UNESCO. World heritage, humanity's gift to future, https://whc.unesco.org/en/activities/487/, 2021
30 Sandbye M. Looking at the family photo album: a resumed theoretical discussion of why and how. Journal of Aesthetics & Culture, 2014, 6 (1): 25419
DOI: 10.3402/jac.v6.25419
31 Ungureanu D, Bogo F, Galliani S, Sama P. Hololens 2 research mode as a tool for computer vision research. 2020
32 Ultralytics. Yolo v5. https://github.com/ultralytics/yolov5, 2021
33 Stacchio L, Angeli A, Lisanti G, Calanca D, Marfia G. Towards a holistic approach to the socio-historical analysis of vernacular photos. ACM Transactions on Multimedia Computing, Communications, and Applications, 2022, 12
DOI: 10.1145/3507918
34 Ma J, Jiang X, Fan A, Jiang J, Yan J. Image matching from handcrafted to deep features: a survey. International Journal of Computer Vision, 2021, 129(1): 23–79
DOI: 10.1007/s11263-020-01359-2
35 Lowe D G. Distinctive image features from scale-invariant keypoints.International Journal of Computer Vision, 2004, 60(2): 91–110
DOI:10.1023/b: visi.0000029664.99615.94

36 Jin H, Wang X, Zhong Z, Hua J. Robust 3D face modeling and reconstruction from frontal and side images. Computer Aided Geometric Design, 2017, 50: 1–13
DOI: 10.1016/j.cagd.2016.11.001

37 itSeez3D Inc. Avatar maker-3D avatar from a single selfie. https://assetstore.unity.com/packages/tools/modeling/, 2022

38 Unity. Unity 2019.4. https://unity3d.com/get-unity/download/archive, 2021

39 Google. Google vr sdk for unity. https://github.com/googlevr/gvr-unity-sdk, 2019

40 Salomoni P, Prandi C, Roccetti M, Casanova L, Marchetti L, Marfia G. Diegetic user interfaces for virtual environments with HMDs: a user experience study with oculus rift. Journal on Multimodal User Interfaces, 2017, 11(2): 173–184
DOI: 10.1007/s12193-016-0236-5

41 Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers, 2003, 35(3): 379–383
DOI: 10.3758/bf03195514

42 Davis F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 1989, 13(3): 319
DOI: 10.2307/249008

43 Taber K S. The use of cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 2018, 48(6): 1273–1296
DOI: 10.1007/s11165-016-9602-2

44 Gerald B. A brief review of independent, dependent and one sample t-test. International Journal of Applied Mathematics and Theoretical Physics, 2018, 4(2): 50
DOI: 10.11648/j.ijamtp.20180402.13

45 Kwak S G, Kim J H. Central limit theorem: the cornerstone of modern statistics. Korean Journal of Anesthesiology, 2017, 70(2): 144
DOI: 10.4097/kjae.2017.70.2.144

46 Stommel M, Katherine J D. Statistics for Advanced Practice Nurses and Health Professionals: Statistics for Advanced Practice Nurses and Health Professionals, 2014

47 Wagner-Menghin M M. Binomial test. Encyclopedia of statistics in behavioral science, 2005

48 ETI. Elettrotecnica imolese. https://www.eti.it/index, 2022

49 Gupta S, Jain S K. A literature review of lean manufacturing. International Journal of Management Science and Engineering Management, 2013, 8(4): 241–249
DOI: 10.1080/17509653.2013.825074

50 Li W, Wu W, Wang H, Cheng X, Chen H, Zhou Z, Ding R. Crowd intelligence in AI 2.0 era. Frontiers of Information Technology & Electronic Engineering, 2017, 18(1): 15–43
DOI: 10.1631/fitee.1601859

51 Niu X, Qin S. Integrating crowd-/service-sourcing into digital twin for advanced manufacturing service innovation. Advanced Engineering Informatics, 2021, 50101422
DOI: 10.1016/j.aei.2021.101422