



# Holograms for seamless integration of remote students in the classroom

Bessie Dominguez-Dager<sup>1</sup> · Francisco Gomez-Donoso<sup>1</sup> · Rosabel Roig-Vila<sup>2</sup> · Felix Escalona<sup>1</sup> · Miguel Cazorla<sup>1</sup>

Received: 10 February 2023 / Accepted: 10 December 2023  
© The Author(s) 2024

## Abstract

The new global scenario imposed by the SARS-CoV-2 virus has given rise to an atypical and problematic situation in multiple spheres. In very little time, the abrupt change from face-to-face to remote has not only required a rapid widespread use of digital technology, but also a change in methodology and communicative interactions. In the field of education, teachers have had to interact in new environments, with the combined use of face-to-face and non-face-to-face teaching being a major challenge. This paper presents the design and implementation of a cyber presence system for educational environments using Microsoft's HoloLens 2 Mixed Reality (MR) headset. A software tool is developed that improves teaching scenarios through communication in mixed environments. The tool enables teachers to integrate the students in the classroom in a common space with remote students connected by videoconference. Our system is not limited to education, however, as it can also be deployed in any setting that requires remote communication, such as companies and governmental institutions.

**Keywords** Remote communication · Education · Mixed reality · HoloLens 2 · Holograms

## 1 Introduction

Despite the advances in society marking a movement toward ubiquity, the new global scenario imposed by the SARS-CoV-2 virus has entailed a process of transition and adaptation from face-to-face to remote environments. In particular, in education, the health crisis forced learning to leave the physical classroom and take place in a virtual space through e-learning (Roig-Vila et al. 2021). Later in the pandemic, when the lockdown was lifted, e-learning was replaced by b-learning with hybrid scenarios (Imants and Van der Wal 2020; Singh 2021).

In this context, teachers have been dealing with the major challenge of simultaneously being in two spaces: face-to-face (classroom) and virtual (videoconference). This situation has generated various difficulties, such as the limitation of movement of what is being transmitted, and the impossibility of seeing all the attendees simultaneously and

perceiving their personal reactions. Therefore, communicative becomes less fluid with a loss of non-verbal communicative information.

In this regard, mixed reality (MR) technologies are vitally important as they allow for the creation of mixed spaces with the integration of virtual environments in physical scenarios. MR has shown great potential to modify the way we interact with the environment and people, enabling representations in the real world of virtual elements through the use of holograms (Ungureanu et al. 2020).

One of the MR interfaces currently being most widely used and adapted to different scenarios is Microsoft HoloLens 2 (Ungureanu et al. 2020). This device presents powerful sensory elements and integrated computational capabilities, such that it allows for the development of novel applications exploiting the fields of MR, artificial intelligence and computer vision (Ungureanu et al. 2020).

In light of the above, this work proposes a hologram-based system to allow remote students to be seamlessly integrated into the classroom. For this purpose, we developed a Microsoft HoloLens 2 MR application that creates a mixed environment for proper interaction and communication between teacher and remote students.

In this regard, our main proposal consists of providing teacher with a tool that enables them to integrate face-to-face

✉ Bessie Dominguez-Dager  
bessie.dominguez@ua.es

<sup>1</sup> University Institute for Computer Research, Universidad de Alicante, Alicante, Spain

<sup>2</sup> Department of General and Specific Didactics, Universidad de Alicante, Alicante, Spain

(students in class- room) with non-face-to-face (students in videoconference). Our proposal can also be deployed in a variety of scenarios that require remote communication, such as in companies or government institutions. Our aim is to prove that the implementation of a cyberpresence system such as that proposed herein improves the effectiveness of communicative interactions between teachers and students and other scenarios.

To further facilitate the teacher learning environment, we consider emotion analysis an important factor, since the larger the number of attendees in a class, the more difficult it is to quickly perceive or evaluate the emotions they are experiencing. Therefore, emotion analysis using computer vision techniques could be a variant to distinguish at the reactions experienced by all the attendees (in classroom and remote). In this way, the teacher may have rapid and direct feedback on how the students are perceiving the class and receiving the content being taught. Thus, they can adapt the content or the way they are being taught in real time or take this into account for future classes.

With the above in mind and to accomplish our main proposal, we designed a web client–server architecture, where the HoloLens 2 application developed accesses remote web services to: (1) remove the background of remote student images captured by their webcam to be able to integrate student holograms into the mixed environment, and (2) classify the predominant student emotions to be shown to the teacher in the MR visualization. Thus, in the proposed implementation the heavy computation part of the system is all carried out in remote services.

Additionally, we focused the work to provide a system that can be used by the students with no economic impact, merely using a simple webcam.

The rest of the paper is organized as follows: First, the state of the art in the field is presented in Sect. 2. Next, Sect. 3 describes the architecture of the proposed cyberpresence system. It also provides an introduction to the HoloLens 2 MR headset and the virtual reality engine that was used to develop the application. In addition, we explain the background removal and emotion classification tasks performed, as well as the interconnection framework designed. Section 4 then presents the methodology used in this work and the technical aspects of the hardware used. Then, Sect. 5 shows the experiments carried out to test the proposed approach and their results. Some limitations of the cyberpresence system are discussed in Sect. 6, while Sect. 7 presents the conclusions of the work.

## 2 Related work

The emergence of MR technologies has been a significant step forward in the way of visualizing and interacting with the real environment and virtual elements. The dynamic

and real-time interaction with digital content provided by MR has been an essential contribution to the development of applications aimed at improving teaching and learning experiences (Tang et al. 2018).

With the emergence of the Microsoft HoloLens interface in 2016, the possibilities of developing applications based on MR with a high level of quality and detail in virtual representations were expanded. Since then, multiple works have been carried out to evaluate and implement MR systems using HoloLens. In the learning context, applications have been highlighted mainly in the field of Medical Science education, specifically in the teaching of anatomy, surgery and orthopedics (Park et al. 2021).

In this sense, researchers have explored MR potential in various studies. For instance, Condino et al. (2018) evaluated MR in hip arthroplasty surgeries, resulting in positive outcomes in workload, performance, visual and audio perception, gesture and voice interactions, and leading to enhanced training speeds for students. Additionally, the Case Western Reserve University School of Medicine developed HoloAnatomy (Wish-Baratz et al. 2019), a holographic anatomy application integrated into their curriculum. The application covers comprehensive anatomy education without the need for real cadavers, utilizing holographic representations of male and female bodies. MR has also been assessed in surgical operation simulations (Paolo et al. 2020) and in the teaching of brain anatomy (Moro et al. 2021), among many other studies.

Furthermore, some systems have been deployed using MR in the sense of “telepresence.” While the above-described approaches used predefined holograms as a complementary learning tool, telepresence with MR aims to use holograms to virtualize and represent users in synchronous communication. In this sense, the term holoportation is used to refer the use of MR to display real-time holograms, facilitating communication between users who are physically distant from one other (Themelis and Sime 2020).

In this regard, one of the telepresence systems deployed with the greatest impact was developed by Orts-Escolano et al. (2016). This system is capable of 360° capture of people, objects and movements within a room, using arrays of depth cameras. The 3D content is modeled, compressed and transmitted to remote participants, with the entire process being performed in real time. In this way, the platform allows for visualizing, listening and interacting between people in different places with close-to-real sensations, as if they were sharing the same physical space.

In order to achieve the speed and computational capacity required for the correct operation of this technology and its real-time execution, several computer systems with graphics processing unit (GPU) have been used. In each recreated capture space, four computers are used for the calculation of the depth and segmentation (one computer for two arrays

of cameras). In addition, the resulting depth maps and the segmented color images are merged and transmitted using another GPU-based computer, requiring high bandwidth communication between the rooms (10 Gigabit Ethernet connection).

Although upgrades such as those proposed in Sanyal and Agrawal (2021) have been made to improve the reconstruction and compression of the models, as well as their transmission, and to reduce the bandwidth, the holoportation system introduced by Orts-Escolano et al. (2016) and its more recent variants (Schmid 2023) continue to require a large amount of high-end hardware to be able to operate.

Other research has employed fewer hardware requirements using different technologies for capturing and reconstructing the 3D hologram of remote users with Microsoft Kinect cameras (Joachimczak et al. 2017). Different approaches have been implemented, representing the remote users as virtual avatars (Piumsomboon et al. 2018) or varying the technology for its visualization in the MR space by using projectors accompanied with lighting effects or with other visual resources (Pejsa et al. 2016; Córdova-Esparza et al. 2018).

For example, in the context of education, Imperial College London developed a system for carrying out remote seminars through the holographic representations of lecturers that appear on stage and deliver real-time teachings to students (Li and Lefevre 2020). This system deploys a first room (remote space) with a specific set-up of a high-definition camera, lighting and audio recording, and a second room with a projector, black curtain backdrop and lighting effects to represent the holograms. The approach achieved a greater motivation among attendees and sense of proximity to face-to-face presence compared to traditional video-conferencing methods.

In the sense of using MR environments to develop education systems that integrate face-to-face (students in the classroom) with non-face-to-face (remote students), all the aforementioned systems shows the possibilities and advantages of telepresence-based communication. However, the use of these approaches is considerably limited by the need for high-definition cameras, depth cameras or multiple arrays as video capturing systems, specific lighting conditions and GPU computers in the room, where the user to be teletransported is located, as well as all the preparation and set-up of the room itself. These requirements make such systems a practically unfeasible solution for everyday use and, therefore, for students at home.

At the same time, systems that use projectors as a visualization tool also require a specific configuration of the environment (light, dark background, etc.), which limits their use in common physical classrooms.

All the above led us to deem necessary a solution that facilitates the implementation of telepresence within the

educational context, with the least number of hardware requirements in a way that allows its use to be increased and provides the possibility of a new way of interacting with reality. Thus, the purpose of this work is to provide the basis for a telepresence system that can be used by multiple users simultaneously, with the lowest possible consumption of hardware resources.

Taking the above into account, we propose an application based on MR and HoloLens 2, which represents a more affordable option for remote students by using a single, simple webcam as the video capturing system. Hence, our proposal does not include the use of 3D cameras or 2D camera arrays, as it would increase the cost of hardware requirements needed for students to use the cyberpresence system.

### 3 Architecture of the system

Figure 1 shows an example of the cyberpresence system developed. It involves a student connected remotely to their webcam, the teacher in the classroom using HoloLens 2 and the students present in the classroom. The webcam image is continuously captured on the student's remote computer and sent to a web server to segment the person (remove background from the image) and classify the predominant emotion.

Meanwhile, the HoloLens that the teacher is using continuously makes requests to the same web server to receive both segmentation and classification information. Each time the HoloLens receives new information, it is processed so that it can be visualized by the teacher in the mixed environment created. Such visualizations (remote student and their emotion) are spawned in a location of the space determined by a 2D image pattern that the teacher places in the classroom, with one pattern for each virtual student. The segmentation and classification tasks are carried out in a GPU-based server system that allows for simultaneous communication between the HoloLens 2 and the remote student's computer.



Fig. 1 Example of the system developed

Next, we describe the components of which the proposed cyberpresence system is composed.

### 3.1 Mixed reality headset

HoloLens 2 is an MR interface created by Microsoft that makes it possible to develop high-comfort immersive experiences for the user. For this purpose, its viewer is specially designed for the use of holograms and it allows gestures to interact with them in a natural way. In addition, it has a sensor for gaze tracking, allowing the interaction to be carried out in an even more natural environment. At the same time, it provides a wide viewing angle and high image resolution (Ungureanu et al. 2020).

In this work, we use the Microsoft HoloLens 2 headset to run an MR application for improving the teaching process in education. In this context, HoloLens 2 is used to represent a mixed environment in which teachers are able to simultaneously visualize students in a physical classroom together with remote students, in the same physical space.

### 3.2 Virtual reality engine

The graphic application created for HoloLens 2 was developed using the Unity3D video game engine, together with the Mixed Reality Toolkit. In the main window of the application, the headset user visualizes the students connected remotely as if they are present in the classroom. To do this, we used Vuforia Engine to detect where to place the virtual students in the MR environment created, using image patterns as targets.

The use of a target in Vuforia means that once it recognizes a pattern previously defined in the application in the real world, through the camera of the MR interface used, virtual content on the position of that target in the mixed environment is displayed. The targets are defined by means of a database created with the patterns (images in this case) we want as a reference to be detected and recognized. For each image in the database, Vuforia applies a set of natural feature detecting algorithms to extract feature points in the image targets. If more features are detected (vertices, lines, areas with high contrast), the images are more highly rated by Vuforia to be considered as good patterns.

Finally, in the MR application, the natural features of the receiving frame from camera are run-time compared with the targets defined in the database. Once a target is detected, Vuforia will track it along the camera's field of view and display the virtual content we have associated with it.

Vuforia also enables a type of tracking that allows virtual content to be incorporated by using a location. The hardware and software used in the MR application then analyze the added content as if it were anchored to a specific location in the real world.

In our case, we used Unity and Vuforia Engine, along with HoloLens spatial mapping and positional tracking systems, to provide tracking of a target even when it is not in view. This means that once a target is detected and recognized, the apparent position in the world of any virtual content whose position is made dependent on that target becomes independent of the headset's location. This results in a stable positioning of the virtual elements, regardless of whether the camera is seeing the defined target or not.

Specifically in the proposed application, each image pattern is placed perpendicular to the HoloLens view, and once the teacher approaches and observes a target, it is detected and recognized. At that moment, two holograms associated with that target, representing one remote student and their emotion, are displayed over the target location. From that moment on, the spatial location of the target is fixed in the HoloLens, so that the holograms associated with it can still be displayed when the target is no longer directly observed by the HoloLens.

For the representation of the remote student in the mixed environment, we create a hologram with the image of the person previously segmented and the background removed. On the other hand, for the representation of the emotion of the student, a hologram is created with an emoji representing the classified emotion. This emoji is placed over the hologram of the projected person and it is updated after each new classification. Three materials were used here, one for each of the classes defined in the emotion recognition task: *Negative*, *Positive* and *Neutral* emotions (Sect. 3.4).

### 3.3 Background removal

The person segmentation is carried out on the images captured by the remote students' webcams to remove the background. This is done with the purpose of guaranteeing a more realistic representation of a person in the mixed environment created. The segmentation was implemented using the SegFormer approach (Enze et al. 2021), which performs efficient semantic segmentation based on transformers and lightweight multilayer perception decoders.

Prior to this selection, we tested other segmentation algorithms such as the mask R-CNN method (He et al. 2017), but, for the proposed system, the SegFormer framework presented the best ratio between execution time and segmentation results. The SegFormer implementation used is fine-tuned on ADE20k dataset. With this implementation, we ran semantic segmentation and then selected the region of the person with the highest area as the final segmentation result.

### 3.4 Emotion classification

To perform the emotion classification task, we used different classification models based on CNN and *K*-NN algorithm

(Mezquita et al. 2020). In addition, one *K*-NN model is presented in which the dimension of the analyzed data is previously reduced. The purpose of using several models was to compare the results of each one, and finally use that which best fitted the MR application.

In general, the classification pipeline developed is defined by two processes: (1) preprocessing of the data to be classified, and (2) emotion classification using one of the implemented models. Figure 2 shows the defined classification structure with the main processes and their sub-processes, as well as the models used.

Our approach leverages a preprocessing stage that is required for the emotion classification. We first perform face detection and then run face alignment. Face detection and alignment implementations will be discussed in Sect. 4. Finally, we carry out a normalization step, resizing the region of interest (RoI) (aligned face) and converting its values to the range [0, 1]. This is mainly done because the CNN model structure used was trained to perform emotion classification on images whose content is fitted to an aligned face with dimensions of 48x48 px. Therefore, the results of this method can be improved by aligning and resizing the face before performing the classification.

The structure of the CNN model, in addition to a set of pre-trained weights for emotion classification, was obtained from Mejia-Escobar et al. (2023). Once the CNN model was created and compiled with the pre-trained weights for emotion classification, we defined a second CNN model from the first one, removing the dense layers and obtaining as output a feature vector of size equal to 4608. This was done to create the *K*-NN+CNN model. The features obtained are the input of the *K*-NN algorithm for performing the classification.

As the feature vector to be used was of considerable size, we defined a variant of the *K*-NN+CNN model, where, after obtaining the features, we performed a dimension reduction using the principal component analysis (PCA) algorithm (Uddin et al. 2021). We called this variant of the model CNN+*K*-NN+PCA.

Finally, it is worth noting that in this task, we use a generalization of the six basic emotions presented by Ekman (1999), and neutral (no emotion expressed) (Barros et al. 2015), into three categories. Based on Akhtar et al. (2019)

and Barros et al. (2015), we defined this generalization as follows: **Neutral**: neutral; **Negative**: sad, fear, disgust, anger; **Positive**: happy, surprise.

### 3.5 Interconnection framework

The interconnection framework is developed as a web communication system that allows information to be exchanged between the HoloLens, the remote students' computer and the system that processes the images. To accomplish this, the client-server architecture in Fig. 3 was implemented. There follows a description of each of its components.

**Webcam client:** This web service is located on the computer of each remote student, capturing images from the webcam continuously and sending them to the main server. The images are encoded as binary strings for sending. Each time this client publishes an image it expects a response with a status that indicates whether it has been correctly received or if an error has occurred.

**Main Server:** Server through which the other four components of the interconnection framework are linked. The main server is continuously receiving the images from remote students' computers (webcam clients). Each time a new image is received, this server sends it to be processed to extract emotion classification and person segmentation information on servers dedicated to performing these tasks, and waits for their responses. If there is more than one webcam client, the images received are processed simultaneously. In all cases, the calls to the two servers are made in parallel by two execution threads. Once the HoloLens

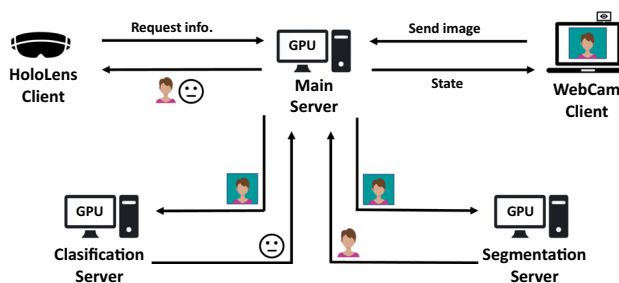
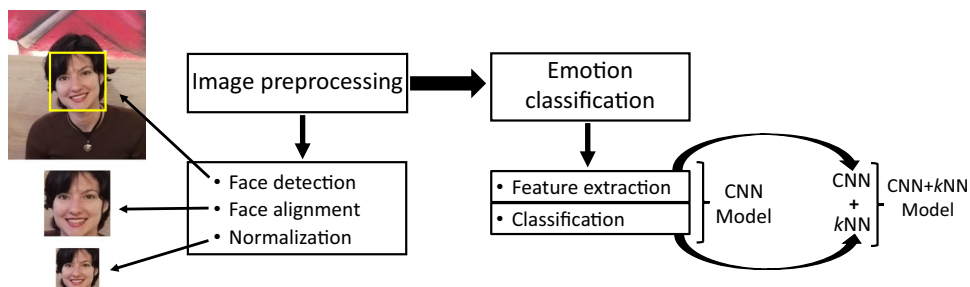


Fig. 3 Interconnection framework design

Fig. 2 Emotion classification process



makes a request to the main server, the last segmentation and classification information received are encoded as a JSON file and returned to the HoloLens. The three servers created were developed on a GPU system to ensure the speed of the system.

**Segmentation Server:** This server runs the segmentation of the person in the images captured by the webcam. First, it decodes the images and performs the necessary processing for segmentation (Sect. 3.3). It then encodes the segmented images without background, as a JSON file and returns it to the main server.

**Classification server:** The emotion classification is performed in this server. To do so, it first decodes the images received from the main server and performs the classification on the main face in the images. The result is indicated by a numerical value corresponding to each defined emotion. Subsequently, the result is encoded as a JSON file and returned to the main server.

**HoloLens client:** The HoloLens client is in charge of requesting the segmented images of the students and their emotions from the main server. Once the requested information is received, it is decoded and can be used in the MR application. This web service was developed in the programming language C# inside the HoloLens application. However, the rest of the services were been developed in Python for convenience in the implementation.

## 4 Methods

### 4.1 Experimental design

With the aim of validating the cyberpresence system developed, a number of experiments were carried out. First, the emotion classification task was quantitatively evaluated to select the classification model that best fits the application. The whole system was then qualitatively analyzed, for which four experiments were carried out.

As we can see, the mixed approach has been used in this research, i.e., a combination of quantitative and qualitative input (Easterbrook et al. 2008). As Matović and Ovesni (2023) point out, both approaches should be linked to the methodological quality of each other. In this sense, we wanted to apply the term complementarity used by Brannen (2005). In this case, qualitative and quantitative results differ per se, but together they generate insights. In this way, the aim has been to achieve the maximum descriptive scope and to provide insight into the characteristics of the system experienced. A non-probabilistic purposive sampling of teachers and students at a university was used to carry out the research. In doing so, a case study design has been applied (Whitmire and Alvin 2019). According to Yin (2009), a case study is “an empirical inquiry that

investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (p. 18). This design does not aim to generalize results, but to understand and analyze a specific situation, in this case, generated by the system created (Tay et al. 2017). In Sect. 5, we present the detailed instruments and procedures for each phase of work in the case at hand.

### 4.2 Technical aspects

For this set of experiments, the web servers were run on the following test set-up: Intel Core i7-6800K with 16 GB of 2400 MHz DDR3 RAM on a MSI X99A SLI PLUS motherboard (X99A chipset). Additionally, the system included an NVIDIA TITAN Xp GPU. The framework of choice was Keras 2.0.8 with Tensor Flow 1.3 as the backbone, running on Ubuntu 18.04. CUDA 8 and cuDNN 7.1 were used to accelerate the computations.

## 5 Results and case studies

### 5.1 Emotion classification experiments

This section describes the tests carried out with the emotion classification models: CNN, CNN+K-NN and CNN+K-NN+PCA, described in Sect. 3.4. The purposes of the tests presented here were:

1. To evaluate the results obtained with different combinations of face detectors, alignment methods and emotion classification models.
2. To select the configuration that presents the best classification option to be used in the MR application with HoloLens 2.

In the preprocessing of the images, the experiments deployed three face detector algorithms: Viola and Jones (V&J) (Agrawal and Khatri 2015), histogram-oriented gradient (HOG) (Dalal and Triggs 2005) and single shoot detector (SSD) (Liu et al. 2016). Additionally, we used two different alignment methods: (1) a first method that we implemented and named self-calculation of face alignment (S-CFA) and (2) face alignment implementation from imutils python module (Rosebroc 2021).

The difference between the two alignment implementations is that the S-CFA method calculates the transformation matrix needed to align the detected face in the input image, and then applies this transformation to that image. Finally, a face detector is rerun on the aligned image to obtain the output RoI. On the other hand, in the imutils implementation, the user defines the desired width and height of the

output RoI, and a value indicating how much of the aligned face will be visible in that RoI.

In order to perform the experiments, the models were trained and tested on the AffectNet database (Ali et al. 2017), except in the case of training the CNN model, where pre-trained weights obtained from Mejia-Escobar et al. (2023) were used.

Here, the AffectNet dataset was chosen for the experimentation because it is the largest facial expression database in world. It is a public database containing more than 1 million images of faces from the Internet. The samples were collected by querying three major online search engines such as, Google, Bing and Yahoo, using 1250 keywords related to different emotions in six different languages (Ali et al. 2017).

For the training and test samples, we redefined these to fit them into neutral, negative and positive, as described in Sect. 3.4. In addition, the samples not corresponding to any of the above were removed from the data set used since they presented no information of interest in the classification pipeline developed.

Subsequently, face detection was performed on the data obtained from the aforementioned process and the images in which no detection occurred were removed. Having detected, aligned and normalized the faces (in both training and test sets), the classification models were trained and then tested.

Table 1 depicts the  $F_1$ -score results in the classification obtained by the different combinations of face detector, alignment method and emotion classification model for the test set. In the table, the number followed by the algorithm  $K$ -NN refers to the number of neighbors used in the experiments with this method.

In Table 1, we can observe that the model using the PCA algorithm yielded the worst results due to the high correlation of the feature vector. On the other hand, the SSD detector also achieved poor classification results since the ROIs produced by this method have a very rectangular shape and therefore suffer great deformation when normalized in the preprocessing stage.

In addition, the table shows that the CNN+ $K$ -NN models yielded a lower  $F_1$ -score compared to the CNN models

across all configurations. This discrepancy can primarily be attributed to the high dimensionality inherent in the feature vector, which requires a more complex function (such as that generated by a CNN) to effectively learn from the data and achieve better results. In the CNN results, we have highlighted in bold the four best  $F_1$ -score values, among which the V&J + S-CFA + CNN configuration obtained the highest score and also produced the best normalized confusion matrix.

Figure 4 shows the normalized confusion matrix generated by the V&J + S-CFA + CNN configuration. The figure shows that this configuration achieves correct classification results for the negative category with values above 0.7. However, certain confusion of neutral and positive emotions with negative is observed. For the positive category, the surprise emotion could be difficult to classify by the model in some cases and, therefore, be confused with a negative emotion.

With the aforementioned results, we selected the V&J + S-CFA + CNN configuration to be used in the application on HoloLens 2 and performed the following experiments.

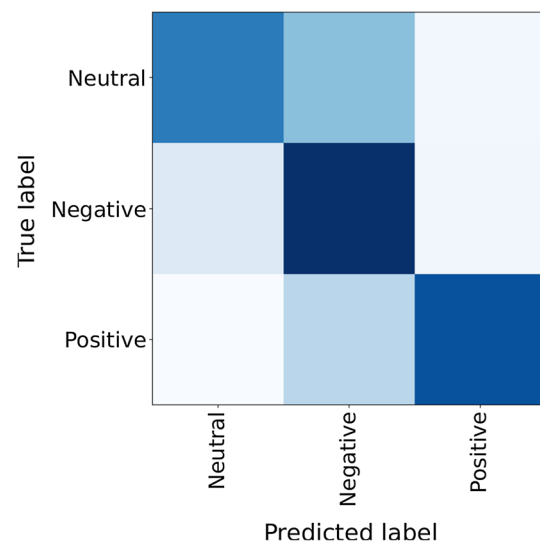


Fig. 4 Normalized confusion matrix produced by the V&J + S-CFA + CNN configuration

**Table 1**  $F_1$ -score results obtained with emotion classification models for different combinations of face detectors and alignment methods

Face Detector	Alignment Method	$F_1$ -score					
		CNN	kNN3	kNN5	kNN7	kNN9	kNN5PCA
V&J	S-CFA	<b>0.6869</b>	0.4932	0.4838	0.4861	0.4699	0.3369
	imutils	<b>0.6409</b>	0.4824	0.4930	0.4990	0.4989	0.2969
HOG	S-CFA	<b>0.6587</b>	0.4803	0.4836	0.4914	0.4736	0.3167
	imutils	<b>0.6568</b>	0.4926	0.5035	0.5106	0.5158	0.3025
SSD	S-CFA	0.5313	0.4046	0.3932	0.3883	0.3697	0.3064
	imutils	0.5055	0.4269	0.4280	0.4196	0.4168	0.2966

## 5.2 Usability of the integrated system

Here, we present the tests performed for the validation of the integrated cyberpresence system. The proposed MR application was qualitatively tested in real scenarios in four different cases. Below, we describe the experiments carried out.

### 5.2.1 Pilot test in an educational environment

In this first test, to simulate the communication between the teacher and a remote student, a validation of the system was performed in a controlled environment using two laboratories of the robotics and tridimensional vision (RoViT) research group of the University of Alicante. In this case, 39 individuals were involved, including nine teachers and 30 students. Table 2 shows demography information of the participants. The experiment consisted of each individual experiencing the use of the proposed system from both the student's and the teacher's side and providing feedback.

The selection of the number of participants for this pilot test, as well as in Sects. 5.2.3 and 5.2.4, conforms to the guidance provided in (Sauro and Lewis 2016). This source advises employing a participant group of approximately 40 individuals to establish the reliability of results in quantitative studies.

Figure 5 shows some examples of visualizations obtained from the HoloLens 2 viewer where correct classification results for the three defined categories of emotions are appreciated, even in situations where the remote student has their hands on their face, is wearing glasses or makes a lateral head movement. Furthermore, correct person segmentation is observed even in the presence of occlusions and gestures. Here, in the implementation of the segmentation algorithm, we used the SegFormer-B0 model, as it provides the fastest performance at the cost of a loss of accuracy compared to higher version models such as SegFormer-B5, as indicated in Enze et al. (2021).

Of all the experiments carried out, the results obtained for the three emotion categories can mostly be said to be

correct. However, some errors in the classification of certain emotions were found, the results for which were consistent with the confusion matrix generated by the configuration selected in Sect. 5.1. Figure 6 shows two examples of incorrect classification results.

However, in some cases, certain segmentation errors are found as shown in Fig. 7. These errors were mainly due to two factors:

1. Situations where the background was characterized by colors that closely resembled those present in the student's skin or clothing, especially in small areas of background contained within a given pose of the person (first image in Fig. 7).
2. Certain gestures, such as raised arm movements, for which the segmentation algorithm does not perform well in all cases (second image in Fig. 7).

After the aforementioned experimentation, an online questionnaire was developed using a Google form as a method of qualitative evaluation of the participants experiences. For the evaluation of the feedback, the five following questions were defined, inspired by the system usability scale (SUS) questionnaire proposed in Brooke (1996) and adapted to the characteristics of our system:

1. How attractive and/or easy to use did you find the designed environment?
2. How would you rate the performance of the interface?
3. What level of realism do you consider present in the application?
4. Do you consider the classification of emotions a positive contribution?
5. How applicable do you consider the proposal presented?

Figure 8 shows the average score per question obtained in the questionnaire. Each question was evaluated in a range from 1 to 5, where 5 corresponds to the best evaluation and 1 the worst. As shown in Fig. 8, the average score obtained

**Table 2** Demography information of participants in the pilot test in an educational environment

Sex	Frequency	Percent (%)
Male	25	64
Female	14	36
Age	Frequency	Percent (%)
20–30 years	25	64
31–40 years	5	13
41–50 years	6	15
51–60 years	3	8
61 years and above	0	0





**Fig. 5** Examples of visualizations obtained from the HoloLens 2 viewer with the proposed cyberpresence system running in the first test. Here, correct person segmentation and emotion classification are shown in each case

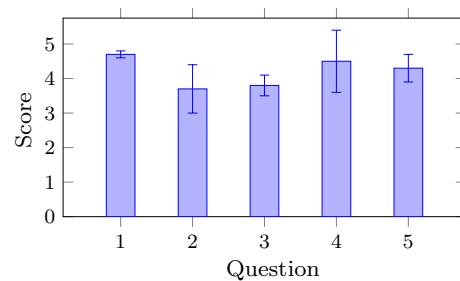


**Fig. 6** Examples of emotion classification errors founded in the first test



**Fig. 7** Examples of segmentation errors founded in the first test

for all the questions was positive, with values starting at 3.7 points. In order to qualitatively evaluate the results of the test, we used an adaptation of the adjectives method proposed in Bangor et al. (2009). This method is based on the association of the numerical score with an adjective scale



**Fig. 8** Average score and standard deviation per question obtained on the online questionnaire in the educational environment. Each question is scored on a range from 1 to 5, where 5 is considered the best evaluation and 1 the worst

(i.e., worst: 0–1.25, poor: 1.25–2.5, ok: 2.5–3.5, good: 3.5–4.0, excellent: 4–4.5, best: 4.5–5.0). This result and participants’ comments indicated that they felt comfortable using the system, giving a qualitative evaluation of above excellent in all aspects.

The highest score, with a qualitative result of best, was obtained in the first question, which indicates that the application succeeded in creating a simple and attractive environment for the test participants. Many users commented that the application provides an easy interaction with the mixed scenario, allowing the teacher to navigate the environment naturally, while visualizing the remote student. At the same time, the interaction between teacher and students less cold

(both face-to-face and remote), contributing positively to the transfer and acquisition of knowledge.

The second indicator with the highest score was the use of emotion classification, with a qualitative evaluation of excellent, as for the rest of the aspects. Most of the participants considered the use of this task an added value to the proposed system. They reported that it allows the teacher to get a better idea of how the students are reacting to the content and/or the teaching method, so that teacher can modify it in real time based on the emotion feedback received.

The applicability of the proposal also scored high. This indicated that the system shown was not only valued as interesting at the level of experimentation, but also its future potential as an application to be implemented in educational environments was also considered.

On the other hand, the indicators of performance and level of realism of the application were rated lowest, although the average scores were still good. The question with the worst rating was that on the level of realism achieved in the application. Here, it is notable that the person's hologram does not represent the whole body. It only shows the part of the body that is captured by the webcam, while the lower part, which should be located below the table, is not displayed, affecting the environment observed.

Although this is a realism limitation, it is a consequence of the student's use of the minimum hardware requirements, such as a simple and single webcam. It is also a result of not requiring previous processes of capturing the whole body before starting the class to save such data and then use it in some form of visualization. All this ensures less time is occupied before starting to use the system. At the same time, it increases the use of the system by students as it does not impose large hardware requirements and, therefore, cost.

Moreover, on certain occasions the level of realism was affected by situations of incorrect segmentation of the person, such as those shown in Fig. 7. This was a drawback that limited the way in which the teacher visualized the remote student at the moment when the incorrect segmentation occurred.

Regarding the performance of the application, although we managed to make it work in real time, the HoloLens has high requirements in the display speed of the frames (60 fps). Therefore, in certain frames, some delay may appear in the update of the holograms with respect to the real state of the student. This does not mean that the application is paralyzed or its representations are slowed down, since if there is no new information, the last information received continues to be displayed, but this information may have a certain delay with respect to the current state of the remote person. This delay is mainly due to the execution time required by the person segmentation and emotion classification algorithms implemented, in addition to the communication times.

Finally, the standard deviation in Fig. 8 shows that most users agreed on their opinions regarding the simplicity of the designed environment, the limitation in the application's realism, and the high applicability of the system. These aspects yielded standard deviation values below 0.5. Nevertheless, there were some disagreements regarding the use of emotion classification. While most participants considered it a positive contribution, some did not feel entirely comfortable with this additional information. Additionally, there was some variety of opinions regarding the system's performance, as most users provided a qualitative evaluation of good, but others were more critical in their evaluation.

### 5.2.2 Full class

In the following case study, a professor gave a full one-hour university class. Here, one student assisted remotely to the class in the same controlled environment as the test described above. In the physical class two students were present. All the participants were aged under 30. The professor was aged over 30. All participants felt comfortable with the system and enjoy its use.

The professor commented that it was a very easy-to-use system, convenient for university environments and guarantees attendance for individuals who for some reason are unable to attend. Moreover, it constitutes an additional contribution for students and teachers because when working through Zoom or another online videoconference platform, the teacher cannot visualize all the students or know whether they are assimilating the information transmitted to them. In addition, the analysis of emotions is also important because it helps to convey how the students are dealing with the content being taught, and based on this, it is possible to adapt the way the class is being taught. Therefore, the professor considered the system a good tool to use in the future in university institutions, but also in meetings, such as in companies when it a more interactive environment is necessary.

Furthermore, the remote student found the system to be an applicable technology, with a good dynamic for individuals that cannot be present in class, making it possible to interact in a more natural and fluid way.

### 5.2.3 Pilot test in a business environment

In this case study, the system was run for business environments. Specifically, 14 entrepreneurship classes were given involving representatives from three companies at three different physical locations. In the test, three individuals participated in each class for a total of 42 participants. Table 3 shows demography information of the participants. The names of the companies involved are not provided for reasons of data protection.

**Table 3** Demography information of participants in the pilot test in a business environment

Sex	Frequency	Percent (%)
Male	30	71
Female	12	29
Age	Frequency	Percent (%)
20–30 years	0	0
31–40 years	16	38
41–50 years	19	45
51–60 years	7	17
61 years and above	0	0

Figure 9 shows three participants in a class of the test, where two of them are remotely connected and the person conducting the class is using the HoloLens 2 in another physical location. The right side of Fig. 9 shows a sample of the resulting visualization in the HoloLens application. Here, we did not run the emotion classification server since we did not consider such information valuable for the test environment.

In this case, the system works in the following sequence: The computer in front of the user in the physical class location is connected to an online meeting continuously recording the person using the HoloLens. At the same time, the remote users are connected to the meeting and can visualize and listen to the person in their respective computers.

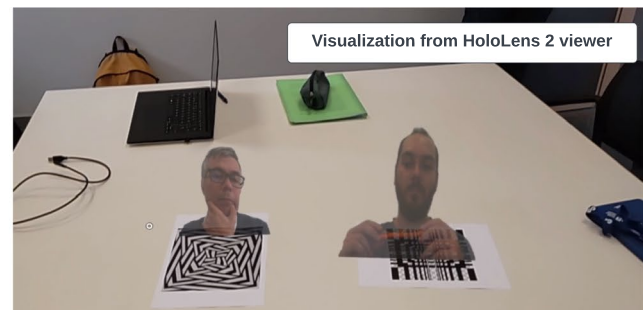
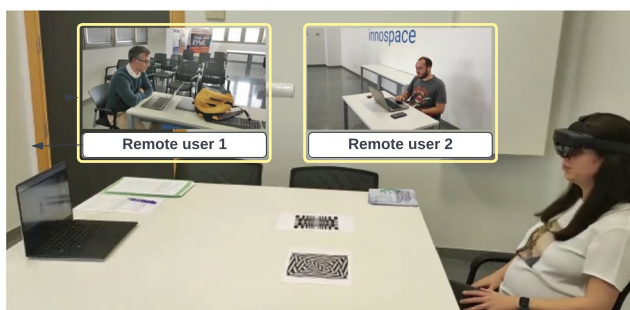
A webcam client runs on each remote user computer and sends information (webcam image captures) to the GPU main server. Through this server, both images are simultaneously processed to extract the person segmentation information from each one. Finally, with the segmentation information obtained, the holograms are created in the HoloLens application. In addition, we added another image pattern for the location of the second remote user in the mixed environment as shown in Fig. 9.

In this case, for the evaluation of the feedback, the technology acceptance model (TAM) proposed by Davis

(1989) was used. This model allows the collection of information to analyze the degree of acceptance of the technology among a given sample in a digital technology scenario. Information is collected on the following dimensions: perceived usefulness, perceived ease of use, perceived enjoyment, attitude toward its use and intention to use.

Taking into account the aforementioned dimensions, a 16-item questionnaire was defined with five response options ranging from “strongly disagree” (1) to “strongly agree” (5). The reliability index of the designed questionnaire was obtained using Cronbach’s Alpha, a procedure recommended by O’Dwyer and Bernauer (2014) for this kind of evaluation. The reliability index value reached was 0.912, which suggests high levels of reliability. The 16 items defined are as follows:

1. I found the system intuitive enough to be used.
2. I think that the system can be employed with autonomy by the user.
3. I think that users of the cyberpresence system will enjoy its use.
4. I found the proposal as a very motivating system for users.
5. I think it is useful to hold meetings with the system.



**Fig. 9** A sample of the system running in a business environment. Here, two remote users participated in an entrepreneurship class given by a person who was using the proposed application in HoloLens 2

6. I think that the system for visualizing remote individuals in a classroom is a good resource for communication.
  7. I found the person visualization system to be a good resource for communication from the remote user side.
  8. I found the prototype to be well thought out.
  9. I think the design of the system is attractive for the user.
  10. I think the options of the system are attractive for the customers.
  11. The options of the prototype are easy and understandable to be implemented widely.
  12. The proposal can satisfy the need for a system to support hybrid teaching sessions.
  13. I would like to use the system in a real situation and put it into practice.
  14. I would introduce the final version of the system in the company or institution plan.
  15. I would recommend using the final version of the system to work with users.
  16. Overall, I rate the proposed system positively.
- As for the attitude toward its use (items 9–12), there is a positive attitude, with a qualitative evaluation between good and best, highlighting the fact that it is a prototype that can be attractive, especially for customers. Additionally, it is also positively valued that the prototype is considered as a support for hybrid communication sessions. However, the prototype's options were assessed as less simple and understandable for widespread implementation.
  - With regard to the intention to use (items 13–15), all the respondents would like to use the system in a real situation and put it into practice, with a qualitative evaluation of excellent. In addition, the use of the system is considered fundamental for working in the company, and the inclusion of the final system version in the business plan was also highly evaluated.
  - Overall (item 16), the proposed system was rated very positively.

Figure 10 shows the average score per question obtained on the 16-item questionnaire in the business environment test, whose qualitative meaning can be interpreted using the adjectives proposed in Sect. 5.2.1. From these results, some conclusions can be drawn:

- Perceived ease of use (items 1–2) was evaluated with a qualitative evaluation of good. This result is consistent with the participants having almost no prior knowledge of the MR interface used in the system.
- As for perceived enjoyment (items 3–4), the values are close to the maximum, with a qualitative result of best. The participants consider that the use of the proposed system provides enjoyment, making it a motivating resource for the business environment.
- Regarding the perceived usefulness (items 5–8), the fact that the system for visualizing remote people in a classroom is a good resource for communication is considered to be of the highest value, with a qualitative result of best. Within this context, the system is considered to be well thought out.

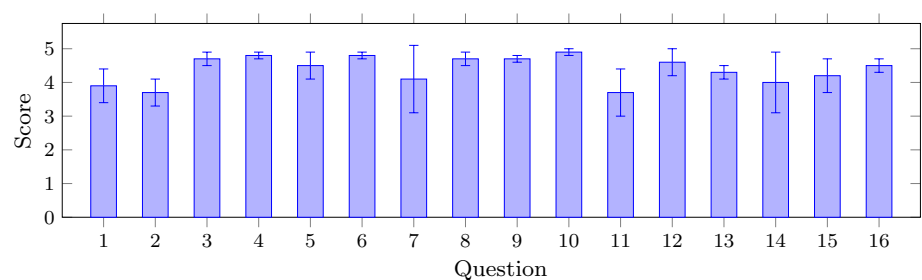
Furthermore, the standard deviation in Fig. 10 shows that users' opinions align on most questions, with deviation values below 0.5. However, some disagreements in opinions were mainly expressed regarding the person visualization system as a good resource for communication from the remote user's perspective (item 7). This is because some participants perceived limitations in the person visualization system's ability to fully convey all the remote user behaviors. Additionally, some participants did not find the prototype easy enough to consider for widespread implementation (item 11). Consequently, there was a variety of opinions regarding the implementation of the final version of the system in the company (item 14).

#### 5.2.4 Pilot test in a local government environment

This case study was carried out in two different city halls in Alicante province, Spain, San Vicente City Hall and La Nucía City Hall. In this case, 12 meetings between personnel from both city halls was carried out. In this test, 36 participants were involved. Table 4 shows demography information of the participants.

Figure 11 shows three of the participants involved in one meeting of the test, as well as an example of the visualization

**Fig. 10** Average score and standard deviation per question obtained by the 16-item questionnaire in the business environment



obtained from the application in HoloLens 2. During this meeting, the mayor of the San Vicente’s city hall tested the system from the HoloLens side. Another employee participated as a remote user from a room in the same building. On the other hand, the first deputy mayor of La Nucía participated as a second remote user from the other location.

In this test, we changed the image patterns used since they produce more robust visualizations than those used in the business environment test. The rest of system configuration and implementation is the same as in the previous test.

In this case, for the evaluation of the feedback, we used the same 16-item questionnaire presented in Sect. 5.2.3.

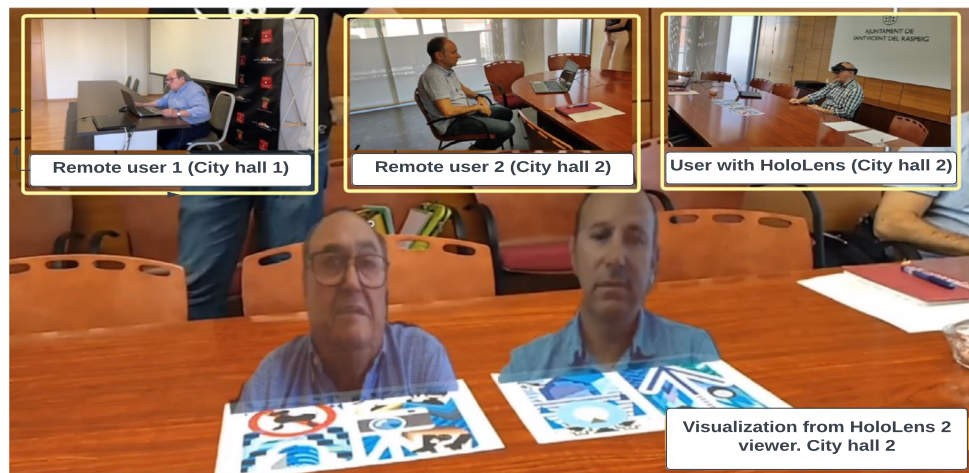
Figure 12 shows the average score obtained per question. This study led to the following conclusions:

- Perceived ease of use (items 1–2) was positively valued in the first item, with an overall score of good, but the value referring to the autonomy of use was scored low, with a qualitative evaluation of only OK, being the lowest rated of the 16 items.
- As for perceived enjoyment (items 3–4), the scores are considered positive, with a qualitative evaluation of excellent.

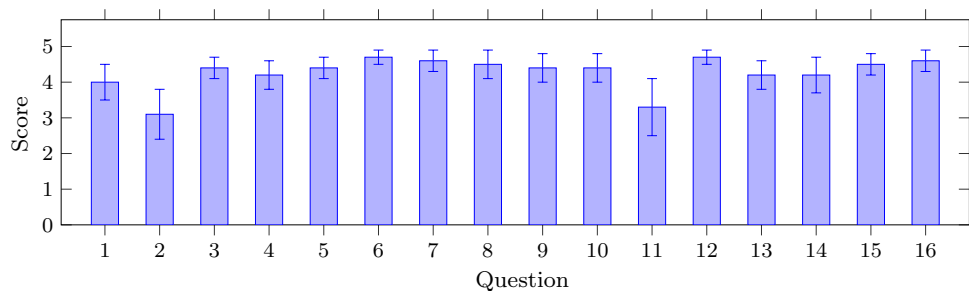
**Table 4** Demography information of participants in the pilot test in a local government environment

Sex	Frequency	Percent (%)
Male	21	58
Female	15	42
Age	Frequency	Percent (%)
20–30 years	3	8
31–40 years	9	25
41–50 years	14	39
51–60 years	7	20
61 years and above	3	8

**Fig. 11** A sample of the system running in a local government environment. Three users were involved at two different city halls of the Alicante province



**Fig. 12** Average score and standard deviation per question obtained by the 16-item questionnaire in the local government environment



- Regarding the perceived usefulness (items 5–8), the scores showed the same trend, always very positive. In general, the system for visualizing remote people in a classroom is considered to be a good resource for communication and the solution proposed by the prototype is well thought out, with an overall evaluation of best.
- As for the attitude toward its use (items 9–12), it was excellent and participants considered that the proposal is attractive for citizens. On the other hand, the fact that the prototype options are easy and understandable to be put into practice in a generalized way is not valued at the same level, and is considered simply OK.
- Regarding the intention to use (items 13–15), the participants positively valued wanting to use the system in a real situation and to put it into practice, as well as considering its use fundamental for working in local government. In addition, including the final version of the system in the institution's plan was rated as excellent.
- Finally (item 16), the proposed system was very positively valued, with a qualitative evaluation of best.

Moreover, when analyzing the standard deviation data in Fig. 12, we can draw some additional conclusions. Most participants' opinions aligned well, with standard deviation values below 0.4. However, discrepancies were observed in the dimension of perceived ease of use (items 1–2). In this regard, participants with more experience in technology found it easier to use the prototype, but for the majority of participants, it proved to be more challenging to use with autonomy. Furthermore, there was a variety of opinions when considering the ease of implementing the prototype on a broader scale. While some participants agreed that it was easy to implement, others expressed some disagreement about it.

## 6 Discussion

Firstly, we must state that overall, the three feedback provided by the users and their qualitative evaluation was positive, as highlighted by the survey. The users' comments also praised the emoji to represent the emotion of the users and the holograms representing them. However, there were some comments regarding the limitations of the current implementation. Some users were concerned about the battery autonomy duration, the field of view and the cost of the device, which are specific issues of the HoloLens 2 headset we used. Flickering and delay in the image transmission were also reported. Additionally, one user stated some discomfort about the fact a pattern is needed wherever the holograms should be displayed. All these matters will be taken into account in future work.

The limitations of HoloLens 2, such as its cost and battery life, have implications for both instructors and students in an MR learning environment.

Firstly, the cost of the device can be a barrier for instructors and educational institutions. The higher price of HoloLens 2 may make it difficult for some instructors to adopt this technology for teaching purposes. However, the cost of the device is expected to fall over time, making it more affordable for universities. Once the cost becomes more reasonable, instructors can leverage the MR environment to enhance their teaching methods and provide students with immersive and interactive experiences.

Secondly, the limited battery life of HoloLens 2 poses challenges for instructors who intend to use the device for extended periods. The two to three hours of active use may not be sufficient to cover a longer class session or multiple consecutive classes. However, the device's compatibility with external batteries provides a solution to this drawback. By utilizing external batteries, instructors can ensure continuous functionality of the device while charging, allowing them to use HoloLens 2 throughout the required class time without interruption.

In terms of the effectiveness and utility of the MR environment based on the learning content, there are implications highlighted in educational environment tests. One of the main limitations is the realism level of the application, as a consequence of using a simple webcam as an image capturing system from the remote students' side. This suggests that the MR environment may not fully replicate the same level of realism and immersion as physical interactions. Instructors should be aware of this limitation when designing learning content and consider alternative approaches or supplementary material to enhance the learning experience of the remote students.

We are aware that other telepresence systems such as some of those mentioned in Sect. 2 (Orts-Escolano et al. 2016; Joachimczak et al. 2017; Sanyal and Agrawal 2021) can achieve a more realistic perception of the MR environment recreated with full-body 3D representations of remote users. However, as we already discussed, these systems involve high hardware requirements with complex video capturing system (high-definition cameras, depth cameras, multi-camera arrays), room set-up and calibration times before starting to use the system. Reasons why they become practically unfeasible solutions for everyday use and students at home. Therefore, although we do not achieve the realism and visualization level accomplish by those systems, we gain in accessibility and affordability of the use of the telepresence system by remote students. At the same time, we allow the easy and immediate use of the system, avoiding calibration and set-up times.

On the other hand, in our application, we found some cases of confusion in the emotion classification task and

errors in person segmentation due to inherent limitations of the segmentation algorithm. These limitations may impact the accuracy and reliability of certain MR applications, such as applications that use emotion recognition as feedback to understand the overall feeling of the class. Instructors should take these limitations into account when designing activities or assessments that heavily depend on such functionalities and consider incorporating alternative methods or tools to achieve the desired learning outcomes.

Overall, while the limitations of HoloLens 2 present challenges, the MR environment still holds promise for instructors and students. It offers opportunities for immersive and interactive learning experiences that can enhance engagement, visualization, and collaboration. Instructors should carefully consider the specific learning goals, content, and activities when incorporating MR into their teaching practices, making informed decisions based on the strengths and limitations of the technology.

In the case studies carried out in the business and local government environments, the general evaluation of the proposed system was very positive, with the exception of the indicator referring to the autonomy of use of the system by the users. Moreover, the fact that the prototype options are easy and understandable to implement in a generalized way was not so positively evaluated either.

Finally, it is important to mention that in the three pilot tests conducted, we concluded that the unequal gender ratio did not influence the results obtained, as no correlations were evaluated with regard to gender. However, we acknowledge this limitation in our study.

## 7 Conclusions and future work

In this work, we present a cyberpresence system using MR to provide the seamless integration of remote students in the classroom. The contributions of this work are multiple. First, a web services system was developed to allow remote communication between teacher and students, which was tested with users in three different physical locations. This represents a novel support to teachers and education, entailing an improvement in situations of training, information and communication in mixed spaces. In addition, the complete system could be tested in business and local government environments.

Second, an emotion recognition system was designed, implemented and evaluated, achieving a  $F_1$ -score value of 0.6869 on a test set of AffectNet dataset. In addition, by using HoloLens 2 headset, a hologram-based cyber presence methodology was developed to represent remote students and their emotions. Finally, participants in the educational environment experiments indicated high levels of satisfaction with the simplicity of the environment created,

the use of the classification of emotions and the applicability of the proposed system. Moreover, in the business and local government environments, the proposed system was positively evaluated.

Despite the promising results, the experiments brought to light some system limitations in terms of the realism level of the application and performance. In addition, certain confusion in the classification of emotions and some limitations in the person segmentation results were found. Furthermore, some disagreement was expressed on the autonomy of use of the system by users, and the ease of the prototype options to be applied in a generalized way, in business and local government environments.

For future works, we plan to further develop the web communication architecture to allow for dynamic interaction with multiple webcam clients. The aim would be to enable the connection of several students at the same time without the main server having predefined the number of students to be connected. We also plan to obtain information from the view of the HoloLens 2 at application execution time, in order to classify the emotions of all the students in the class (face-to-face and remote students). Moreover, we intend to improve the classification results. Finally, we aim to project the representation of remote students with a method that does not require the use of an image pattern. We also plan to improve the autonomy of use of the system.

Finally, it is worth noting that this study was conducted under the supervision of the ethics committee of the University of Alicante, and all the participants gave their informed consent. No datasets, as in a large set of labeled data or images, were generated, although we collected the participants' responses to the surveys, which will be available under request. As mentioned, the participants granted their permission and were informed about the study and how their data would be handled. Regarding the ethical implications and the concerns about personal data, it should be stated that the identity, images, biometric data and any other kind of data collected from the user during the deployment of the system is properly codified and encrypted using a public-key methodology, and so it is unlikely that in the event of someone capturing the data, they could browse the actual content. In addition, the images and other data are processed by the different algorithms in a secure local server in an autonomous fashion, and so no humans interfere in the process. No personal data are stored whatsoever; only the predictions of the system are saved. It is impossible to either identify any user or reconstruct their data from the predictions.

**Acknowledgements** This work was carried out under the framework of the grant PID2022-138453OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". Partial work carried out by Bessie Dominguez-Dager has been funded by the grant UAIND21-04C from University of Alicante.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

**Availability of data and materials** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors declare that the research was conducted without conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agrawal Sa, Khatri P (2015) Facial expression detection techniques: based on viola and jones algorithm and principal component analysis. In: 2015 fifth international conference on advanced computing and communication technologies. IEEE. pp 108–112
- Akhtar MS, Chauhan DS, Ghosal D, Poria S, Ekbal A, Bhattacharyya P (2019) Multi-task learning for multi-modal emotion recognition and sentiment analysis. arXiv preprint [arXiv:1905.05812](https://arxiv.org/abs/1905.05812), <https://doi.org/10.48550/arXiv.1905.05812>
- Ali M, Behzad H, Mahoor MH (2017) Affectnet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans Affect Comput 10(1):18–31. <https://doi.org/10.1109/TAFFC.2017.2740923>
- Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 4(3):114–123
- Barros P, Weber C, Wermter S (2015) Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction. In: 2015 IEEE-RAS 15th international conference on humanoid robots (humanoids), pp 582–587. IEEE <https://doi.org/10.1109/HUMANOIDS.2015.7363421>
- Brannen J (2005) Mixing methods: the entry of qualitative and quantitative approaches into the research process. Int J Soc Res Methodol 8(3):173–184. <https://doi.org/10.1080/13645570500154642>
- Brooke J et al (1996) SUS—a quick and dirty usability scale. Usability Eval Ind 189(194):4–7
- Condino S, Turini G, Parchi PD, Vigliani RM, Piolanti N, Gesi M, Ferrari M, Ferrari V (2018) How to build a patient-specific hybrid simulator for orthopaedic open surgery: benefits and limits of mixed-reality using the microsoft hololens. J Healthc Eng. <https://doi.org/10.1155/2018/5435097>
- Córdova-Esparza D-M, Terven J, Jiménez-Hernández H, Herrera-Navarro A, Vazquez-Cervantes A, García-Huerta J-M (2018) Telepresence system based on simulated holographic display. arXiv preprint [arXiv:1804.02343](https://arxiv.org/abs/1804.02343)
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, vol 1, pp 886–893 <https://doi.org/10.1109/CVPR.2005.177>
- Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, pp 319–340. <https://doi.org/10.2307/249008>
- Easterbrook S, Singer J, Storey M-A, Damian D(2008) Selecting empirical methods for software engineering research. In: Guide to advanced empirical software engineering, pp 285–311
- Ekman P et al (1999) Basic emotions. Handbook Cogn Emot 98(45–60):16
- Enze X, Wang Wenhai Yu, Zhiding AA, Alvarez Jose M, Ping L (2021) Segformer: simple and efficient design for semantic segmentation with transformers. Adv Neural Inf Process Syst 34:12077–12090
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-CNN. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969
- Imants J, Van der Wal MM (2020) A model of teacher agency in professional development and school reform. J Curric Stud 52(1):1–14
- Joachimczak M, Liu J, Ando H (2017) Real-time mixed-reality telepresence via 3d reconstruction with hololens and commodity depth sensors. In: Proceedings of the 19th ACM international conference on multimodal interaction, pp 514–515
- Li N, Lefevre D (2020) Holographic teaching presence: participant experiences of interactive synchronous seminars delivered via holographic videoconferencing. Res Learn Technol. <https://doi.org/10.25304/rlt.v28.2265>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) SSD: single shot multibox detector. In: European conference on computer vision. Springer, Berlin, pp 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- Matović N, Ovesni K (2023) Interaction of quantitative and qualitative methodology in mixed methods research: integration and/or combination. Int J Soc Res Methodol 26(1):51–65
- Mejia-Escobar C, Cazorla M, Martinez-Martin E (2023) Towards a better performance in facial expression recognition: a data-centric approach. Comput Intell Neurosci (**in press**)
- Mezquita Y, Alonso RS, Casado-Vara R, Prieto J, Corchado JM (2020) A review of k-NN algorithm based on classical and quantum machine learning. In: International symposium on distributed computing and artificial intelligence. Springer, Berlin, pp 189–198
- Moro C, Phelps C, Redmond P, Stromberga Z (2021) Hololens and mobile augmented reality in medical and health science education: a randomised controlled trial. Br J Edu Technol 52(2):680–694. <https://doi.org/10.1111/bjet.13049>
- O'Dwyer LM, Bernauer JA (2014) Quantitative research for the qualitative researcher. SAGE Publications, Thousand Oaks. <https://doi.org/10.4135/9781506335674>
- Orts-Escolano S, Rhemann C, Fanello S, Chang W, Kowdle A, Degtyarev Y, Kim D, Davidson PL, Khamis S, Dou M et al (2016) Holoportation: virtual 3d teleportation in real-time. In: Proceedings of the 29th annual symposium on user interface software and technology, pp 741–754
- Paolo C, Alfredo L, Alessandro C, Maria SG, Giampiero D, Francesco O (2020) An interactive real-time cutting technique for 3d models in mixed reality. Technologies 8(2):23. <https://doi.org/10.3390/technologies8020023>
- Park S, Bokijonov S, Choi Y (2021) Review of microsoft hololens applications over the past five years. Appl Sci 11(16):7259. <https://doi.org/10.3390/app11167259>
- Pejsa T, Kantor J, Benko H, Ofek E, Wilson A (2016) Room2room: Enabling life-size telepresence in a projected augmented reality environment. In: Proceedings of the 19th ACM conference on computer-supported cooperative work and social computing, pp 1716–1725



- Piumsomboon T, Lee GA, Hart JD, Ens B, Lindeman RW, Thomas BH, Billinghamurst M (2018) Mini-me: an adaptive avatar for mixed reality remote collaboration. In: Proceedings of the 2018 CHI conference on human factors in computing systems, pp 1–13
- Roig-Vila R, Solano MEU, Merma-Molina G et al (2021) La comunicación en el aula universitaria en el contexto del covid-19 a partir de la videoconferencia con google meet. *Revista Iberoamericana de Educación a Distancia (RIED)* 24(1):197–220. <https://doi.org/10.5944/ried.24.1.27519>
- Rosebrock A (2021) Github - pyimagesearch/imutils. <https://github.com/jrosebr1/imutils>
- Sanyal H, Agrawal R (2021) Study of holoportation: using network errors for improving accuracy and efficiency. In: Proceedings of international conference on sustainable expert systems. Springer, Berlin, pp 107–118 [https://doi.org/10.1007/978-981-33-4355-9\\_9](https://doi.org/10.1007/978-981-33-4355-9_9)
- Sauro J, Lewis JR (2016) Quantifying the user experience: practical statistics for user research. Morgan Kaufmann, Burlington
- Schmid III Maj. Gen. (Dr.) Josef (2023) Challenges and opportunities in space medicine, space flight, and holoportation. *Telehealth Med Today* 8(2), Jan 2023 <https://doi.org/10.30953/thmt.v8.385>. <https://telehealthandmedicinetoday.com/index.php/journal/article/view/385>
- Singh H (2021) Building effective blended learning programs. In: Challenges and opportunities for the global implementation of E-learning frameworks. IGI Global, pp 15–23
- Tang YM, Au KM, Leung Y (2018) Comprehending products with mixed reality: geometric relationships and creativity. *Int J Eng Bus Manag.* <https://doi.org/10.1177/1847979018809599>
- Tay LY, Melwani M, Ong JL, Ng KR (2017) A case study of designing technology-enhanced learning in an elementary school in Singapore. *Learn Res Pract* 3(2):98–113
- Themelis C, Sime J-A(2020) From video-conferencing to holoportation and haptics: how emerging technologies can enhance presence in online education? *Emerging technologies and pedagogies in the curriculum*, pp 261–276
- Uddin MP, Mamun MA, Hossain MA (2021) PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech Rev* 38(4):377–396
- Ungureanu D, Bogo F, Galliani S, Sama P, Duan X, Meekhof C, Stühmer J, Cashman TJ, Tekin B, Schönberger JL et al (2020) HoloLens 2 research mode as a tool for computer vision research. arXiv preprint [arXiv:2008.11239](https://arxiv.org/abs/2008.11239)
- Whitmire D, Alvin C (2019) A case study in software testing: verification of a face identification algorithm for planar graphs. *J Comput Sci Coll* 35(3):173–184
- Wish-Baratz S, Gubatina AP, Enterline R, Griswold MA (2019) A new supplement to gross anatomy dissection: holoanatomy. *Med Educ* 53(5):522–523. <https://doi.org/10.1111/medu.13845>
- Yin RK (2009) Case study research: design and methods, vol 5. SAGE, Thousand Oaks

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.