*Review*

# Site Experience Enhancement and Perspective in Cultural Heritage Fruition—A Survey on New Technologies and Methodologies Based on a "Four-Pillars" Approach

Agnese Augello *,†,‡ , Ignazio Infantino †,‡ , Giovanni Pilato †,‡ and Gianpaolo Vitale †,‡

Institute for High Performance Computing and Networking (ICAR)—National Research Council of Italy (CNR), Via Ugo La Malfa 153, 90146 Palermo, Italy; ignazio.infantino@icar.cnr.it (I.I.); giovanni.pilato@icar.cnr.it (G.P.); gianpaolo.vitale@icar.cnr.it (G.V.)

*   Correspondence: agnese.augello@icar.cnr.it
†   Current address: Via Ugo La Malfa 153, 90146 Palermo, Italy.
‡   These authors contributed equally to this work.

**Abstract:** This paper deals with innovative fruition modalities of cultural heritage sites. Based on two ongoing experiments, four pillars are considered, that is, *User Localization*, *Multimodal Interaction*, *User Understanding* and *Gamification*. A survey of the existing literature regarding one or more issues related to the four pillars is proposed. It aims to put in evidence the exploitation of these contributions to cultural heritage. It is discussed how a cultural site can be enriched, extended and transformed into an intelligent multimodal environment in this perspective. This new augmented environment can focus on the visitor, analyze his activity and behavior, and make his experience more satisfying, fulfilling and unique. After an in-depth overview of the existing technologies and methodologies for the fruition of cultural interest sites, the two experiments are described in detail and the authors' vision of the future is proposed.

**Keywords:** cultural heritage; localization; multimodal interaction; augmented reality; affective computing; gamification; cognitive architectures

## 1. Introduction

The modalities for the fruition of cultural heritage sites are continuously evolving due to the availability of increasingly innovative technologies, advanced data analysis methodologies and, in general, the growing appeal of innovative sightseeing approaches for tourists.

It is well known that the optimal fruition of cultural heritages requires a difficult trade-off between the different needs of a broad and variegated set of visitors and the peculiarities of the specific site. In fact, many objects cannot be touched or seen from a close distance because of restrictions that aim to avoid deterioration. On the other hand, some detailed information or specific details can be focused on the visitor's interests. In this context, innovative technologies and methodologies can be exploited in the cultural heritage sector, ensuring straightforward accessibility and simplifying the definition of installations and routes since most of them are already available on portable smart devices.

Besides, a specific type of equipment can be chosen to comply with the particular characteristics of the environment [1]. As an example, indoor or outdoor location can lead to different choices to manage the user's localization; furthermore, the structure of the environment could limit the use of specific technologies. The variety of approaches and methodologies can enable the design of customized and involving fruition modalities, improving the compliance to the visitors' needs, for example, the design of group paths based on gamified approaches for a younger target or the design of individual tracks properly conceived to respond to specific needs (as in the case of people with disabilities).

The possibility of using wearable and not invasive devices allows for acquiring data of interest, like the location, the current routes, and the visitor's visual attention.

Among new facilities, Virtual Reality (VR), Augmented Reality (AR) and natural language interaction systems allow for enriching and extending the visiting experience [2]. Even if the facilities are tied to technologies, both VR and AR can be defined, avoiding specifying particular hardware as in [3] where VR is defined "a real or simulated environment in which a perceiver experiences telepresence", and in [4] where AR is defined as a technique to show extra information over the real world, respectively. Between VR and AR, there is a "virtual continuum" where mixed reality encompasses everything between reality and a virtual environment [5].

VR and AR have been attracting researchers' interest for several years; however, since their spread depends on several factors, such as technologies, society's needs and applications, the research trend showed peaks and stagnations [6].

On the other hand, these technologies and related applications play a crucial role in widening the potential market. It can be noted that AR and VR were not devised for cultural heritage, in fact, starting from games applications, both VR and AR have been driven by the health sector, education and industry while technologies benefit from the increased computing capacity and the lowering of the devices [7], particularly smartphones [8]. Recently VR and AR, and related technologies have also been adopted in cultural heritage to improve the site's experiences. Nowadays, the wide availability of smart devices with both affordability and high computational power, makes their application very widespread, and relatively inexpensive. Moreover, the use of intelligent data analysis and artificial intelligence, to process the acquired information, makes it possible to effectively profile the visitor, who can consequently receive personalized and non-trivial suggestions and recommendations to continue the visit. These advancements make it possible not only to focus on the visitor but also to transform the cultural site into an intelligent multimodal environment, emphasizing the emotional aspects of the interaction.

In our vision, we conceive the cultural heritage site as an intelligent entity embodied in a multimodal interactive environment. This entity can perceive human beings' activities and their feelings, reacting accordingly by talking or by giving sounds, images, and colors as feedback. Such an entity's behavior should be driven by the desire of establishing a social, emotional, engaging connection with the visitors while supporting them in the site fruition. Therefore, in this work, we analyze the technologies and the methodologies proposed in the last years in the context of cultural heritage concerning four main aspects that we consider as the pillars of our vision: (a) *User Localization*: the user must be localized in the environment, the knowledge of his position and of the objects surrounding him enables the possibility of providing a contextualized and personalized information; (b) *Multimodal Interaction*: the user should have the possibility of interacting in a variety of ways, such as through spoken language, gestures and glances; (c) *User Understanding*: understanding how people make use, interact, and catch the essential elements of a cultural heritage site allows improving their knowledge and comprehension; and (d) *Gamification*: a closer link between the visitor and the environment can be created by playing serious games supporting cultural heritage learning. It should be underlined that all the four pillars aim to enforce the overall interaction and interest between users and cultural heritage.

In the following sections, after positioning our paper compared to the state-of-the-art in Section 2, we provide an overview of the main systems dealing with the aspects mentioned above by explaining why they are particularly relevant for our vision. In particular, in Section 3, we tackle user localization both in indoor and outdoor spaces by summarizing how it is possible to sense the user's presence in the environment and obtain its position. In Section 4, we outline the existing technologies to perceive and interact with the museum visitors through different verbal and not verbal modalities. In Section 5, we deepen the methodologies used to make deductions about the user's profile, needs and emotions. In Section 6, we give an overview of the gamification approaches currently used in cultural heritage to increase visitors' attention and engagement, which could inspire

the formalization of strategies for our museum intelligent agent. Finally, in Section 7, we provide some concluding remarks about the survey and a description of our vision, inspired by technologies and solutions identified in the literature. References are then categorized by an overview table based on their relevance with one or more of the four pillars we focused in the paper.

## 2. Position of This Work Compared to the Current State of the Art

The approach proposed in this paper is based on the four pillars, namely: User Localization, Multimodal Interaction, User Understanding and Gamification, since, as we discussed in the Introduction, we conceive cultural heritage as an interactive intelligent environment able to localize and identify the user and interact with him engagingly and amusingly. The literature proposes many papers dealing with the improvement of cultural heritage fruition by innovative technologies; however, most of them consider an update of existing solutions employing one of the four pillars proposed here. We have observed great interest in augmented and virtual reality technologies; several surveys, such as [2,9], focus on this specific technological aspect, highlighting its importance to ensure both amusement and accessibility (especially when physical access is constrained), discussing future research directions and comparing different categories of immersive reality. Other surveys focus on the methodologies of machine learning more exploited in cultural heritage [10], provide an overview of existing 3D repositories [11], or outline museum gaming technologies and applications [12,13]. There is a review of interactive systems for cultural heritage [14], but the works are analyzed from the perspective of their empirical evaluation, that is, the degree to which the system satisfies user goals and expectations.

To the best of our knowledge, no other survey analyzed the Cultural Heritage literature considering all together these four aspects that in our opinion are altogether significant to develop an intelligent and engaging fruition to cultural heritage.

Moreover, a discussion of one of these pillars, inserted in a broader discussion including the other three, in our opinion, can be of greater inspiration for a more effective, multi-faceted approach. As an example, a reader interested in Gamification could be inspired by the discussion about the other pillars, considering some of the reviewed technologies and methodologies for defining new gaming strategies. We believe that a four-pillars vision contributes to covering, from a holistic perspective, the main issues of cultural heritage fruition—user localization, accessibility through different modalities of interaction, adaptation to user profile and interests, and the provision of an amusing experience. The reason for doing this review came from some of our ongoing experiences in the field of innovative content fruition and, in particular, in the context of cultural heritage (the experiences will be discussed in Section 7). We aimed to address an overview of the state-of-the-art and, in consideration of the fact that rapidly evolving technologies can be leveraged in this context, we were unable to find a sufficiently recent work that was useful for our purposes.

We concluded that a more updated review was important. Just to give an example, an important aspect in museum navigation and in contextualized information provision is that of localization, an area where significant changes are observed over a short period. Many interesting technologies are emerging and evolving separately, we assumed that this work would be of interest to those like us who wanted to develop something dedicated in the field, figuring out how to bring different technologies and methodologies together.

We planned this review by working according to a well-established procedure. The first step was to analyze the main issues to be addressed in this context, leading to the identification of the four pillars mentioned above. We then used the most important research repositories to consider.

We analyzed works between 2017–2021, mainly concerning the more methodological aspects, with some exceptions for a few approaches that, although older, we found inspirational, and for some technological aspects where we considered a reduced temporal range.

After a first selection and review of the collected works, we outlined those more meaningful for the four pillars. Finally, we summarized the most inspiring contributions for our vision, in Section 7, by means of a concise table.

### 3. Perception of the User in the Environment: Methodologies and Technologies for Localization

To provide any service to the user, he/she must be located in the context. There are different methods and technologies in this regard that differ in various parameters, including accuracy, cost, energy consumption; this area is constantly evolving thanks to the use of mobile devices owned by the users. This section analyzes the most recent technologies and methods for localization, considering also the contribution of those related to the internet of things.

#### 3.1. Methodologies for Indoor Localization

Indoor localization, meaning the process to retrieve a user or device localization in an indoor setting or environment, has been boosted in recent years by smartphones and wireless devices. Indoor device localization has been used, starting from a few decades in industrial settings and robotics, however, only less than a decade the possibility to track a user enabled a wide range of applications and services. Also, the Internet of Things (IoT) approach based on the connection of end-to-end billion devices, starting from long-range technologies, has been extended to short-range localization even if in cooperation with different communication interfaces [1]. In general, indoor localization requires a finer granularity compared to outdoor localization meaning a more significant information on a smaller area; this is the reason why it encompasses different localization techniques, technologies and systems. Three different localization approaches can be recognized. In the Device Based Localization (DBL), the user position is retrieved based on the reference nodes ; it is applied where assistance in navigating around space is required by the user. Differently, in the Monitor Based Localization (MBL), the reference nodes passively obtain the user's position; the user is often tracked to provide services. Finally, proximity detection aims to estimate a distance between the user and a Point of Interest (PoI). Proximity detection is already used in a shopping mall, as an example, where the user can be informed on promotions based on its position. It is also of interest since in a museum or in the context of the fruition of cultural goods, a series of additional multimedia information can be enabled for the user when approaching an object [1]. As concerns the indoor localization techniques, one of the most straightforward and widely adopted approach is the received signal strength (RSS). It exploits the power of the signal at the receiver to estimate the distance between the receiver and the transmitter. The DBL requires more reference points; then trigonometry is applied to calculate the distance. In the MBL case, the strength of the signal at a reference point is used for calculating the position of the user. RSS proximity services can use a single reference node to build a geofence to estimate the proximity. The RSS is simple and cheap, but it can exhibit poor accuracy due to signal attenuation and noise; it can be improved by implementing more complex algorithms [15,16]. The Channel State Information (CSI) techniques composed of the Channel Impulse Response (CIR) and of the Channel Frequency response (CFR) are conceptually similar to RSS but show a finer granularity being based on the detection of both amplitude and phase of the signal. Many cheap network interface controllers (NIC) IEEE 802.11 NICs cards can provide more information resulting in stable measurements and higher localization accuracy [15]. In Scene Analysis techniques, measurements are taken offline (using RSS or CSI), then they are compared to the online ones. The comparison is performed by probabilistic methods, artificial neural networks, k-nearest neighbor or support vector machines [17]. The Angle of Arrival (AoA) method adopts antennas array and calculates the time difference at the individual antenna. In general, this method is quite accurate but in comparison with RSS, it requires more complex hardware and precise calibration [18]. The Time of Flight (ToF) or Time of Arrival (ToA) uses the propagation time of the signal to compute the distance between the receiver and the transmitter; it is

suitable for both MBL and DBL scenarios. On the other hand, a strict synchronization is required, and estimation accuracy depends on the sampling rate and signal bandwidth. It should be underlined that in case of unavailability of a sight path, direct line localization error remains [19]. Time Difference of Arrival (TDoA) employs more transmitter (three at least) and calculates the differences of the distances with the receiver. In this case, strict synchronization is still required but only among transmitters [19]. The Return Time of Flight (RToF) is based on the measurement of the round trip propagation time meaning that it considers the path transmitter-receiver-transmitter. In principle, it is similar to ToF but a less strict synchronization is required. On the other hand, sampling rate and bandwidth are more critical since the signal is transmitted and received twice [19]. The Phase of Arrival (PoA) is based on the phase difference between the transmitter and the receiver to compute the distance. Localization is performed by TDoA based algorithms. This method can be used RSSI, ToF, TDoA to improve the localization accuracy [20,21]. As concerns disadvantages, it requires line-of-sight for high accuracy.

### 3.2. Technologies for Localization

Some technologies providing localization services are already consolidated on the market, such as WiFi, Bluetooth, ZigBee, and UltraWide Band (UWB); in addition, emerging technologies, like SigFox, LoRa, IEEE 802.11ah and Weightless, are gaining interest. The IEEE 802.11 standard, known also as WiFi, was conceived for providing network capabilities and Internet connection to different devices. The initial reception range was about 100 m increased to about 1 km by the IEEE 802.11ah. The aforementioned RSS, CSI, ToF and AoA techniques (alone or in combined form) allow WiFi-based localization systems [22–24]. The latest version of Bluetooth (Bluetooth Low Energy—BLE), can be used with RSSI, AoA, and ToF for localization. Two BLE based protocols have been proposed for proximity-based services—iBeacons proposed by Apple and Eddystone proposed by Google. IBeacons transmit signals (beacon) at periodic intervals, by RSSI the proximity is classified into immediate (<1 m) near( 1–3 m), far (>3 m) and unknown. Zigbee is based on IEEE 802.15.4 standard, it defines the higher levels of the protocol stack, and it is exploited in wireless sensor networks (WSN). Although it is favorable for localization of sensors in WSN, it is not readily available on the majority of the user devices, as a consequence, it would be preferable for users to use a personal device such as a smartphone both for pandemic and for cheapness,  it is considered not suitable for indoor localization [25]. The Ultra Wide Band (UWB) technology is based on ultra short-pulses with a time period lower than 1 ns. It is attractive for indoor localization since it is relatively immune to interferences due to other signals and exhibit a low power consumption; despite this, it has limited use in consumer products and portable user devices [26,27]. The Visible Light Communication (VLC) uses light sensors by LEDs behaving like iBeacons; on the other hand, a sight line is necessary between LED and sensors [28,29]. The acoustic signal technique shows a potential advantage since it can exploit the smartphones sensors to receive signals by sources and can use the ToF technique. However, the low emission level to be respected to avoid sound pollution and the energy consumption limit the use. The problem of sound pollution is solved by ultrasound-based localization but limitations due to the degradation of the signal with humidity and temperature (contrarily to RF) remain [30].

### 3.3. Localization by Emerging IoT Technologies

The traditional indoor location-based services are often associated with people positioning and tracking by exploiting the wireless signal transmitted from their personal devices; for this reason, they are becoming a more and more integral part of the wider Internet of Things (IoT) paradigm whose related technologies are of interest also for proximity detection and localization. In this perspective, the idea of connecting different "things" can encompass localization, enhancing the wide range of provided services. IoT are considered one of the six "disruptive civil technologies" by the US National Intelligence Council (NIC) [31]; in addition to expected services related, as for example, to security, marketing,

health and so on, the fruition of cultural heritage can achieve attractive benefits taking into account also that a wide spreading on the market implies a decreasing of costs. The main emerging radio technologies conceived for IoT and related to localization are SigFox, LoRa, IEEE 802.11ah, and Weightless [1]. They all characterized by a wide reception range in outdoor applications (from 1 km of IEEE 802.11ah to 50 km of SigFox), and extremely low energy consumption. As concerns the indoor applications, SIgFox, LoRa and Weightless suffer of signal attenuation due to the walls implying a relevant outdoor-indoor attenuation whereas IEEE 802.11ah due to its shorter-range network architecture and to significantly lower propagation loss through free space and walls/obstructions can be expected a good candidate for indoor localization [1].

## 4. Multimodal Interaction

In the context of cultural heritage, it is important to ensure a high degree of accessibility. The possibility of offering new and different modalities of interaction provides a complete fruition experience when some assets cannot be seen from a close distance or in their details or even touched by the user (for people safety, artworks protection, or accessibility issues). This is particularly important for people who, for some impairments, may have difficulty in visiting the site in the traditional way. Interactive interfaces can be tangible, collaborative, device-based, sensor-based, hybrid, and multimodal [2]. Starting from the observation that all of our senses and their cross modal effects contribute in our perception and understanding of an environment, the paper [32] underlines the importance of a multisensory reconstruction of a past heritage environment. That work discusses systems reconstructing past heritage environments by exploiting variables such as temperature changes, sounds, tactile sensations, visual elements.

As discussed in the introduction, we conceive the cultural heritage site as an intelligent entity, interacting with the visitors with different input/output modalities. In the following subsections we analyze multimodal systems by collecting them according to the predominantly addressed modality of interaction. In particular, we have considered three sensory input/output channels that it is currently possible to recognize in a typical human-computer interaction: *Auditory*, *Visual* and *Tactile*, extending them with an *Artistic* channel. In this last group we placed mainly interactive exhibits capable of perceiving, in some way, the presence of the visitor ( his movements, his heartbeat, activities, etc.), and reacting by expressing some sort of internal mood with artistic elements (such as sounds, colours, pictures).

### 4.1. Interaction Based on the Auditory Channel

One modality consists of using sounds or pre-recorded texts, where in some cases the speech can be automatically and dynamically generated by a text-to-speech engine. Often it is proposed a spoken natural language interface as a usable way of interaction. Such a vocal interface requires a robust speech recognition system and the analysis and comprehension of the recognized text. However, it can give inaccurate results in overcrowded environments and therefore demands specific requirements, such as the use of accurate microphones. The MAGA system exploits a chatbot interface and radio frequency identification (RFID) tags attached to points of interest [33]. Users used a PDA running the MAGA client application able to process the data of RFID tags. Every time a tag is detected, the RFID Module communicates the position to a chatbot, that, as a consequence, starts the interaction providing information about the detected item and waiting for user requests. An application for character-based guided tours on mobile devices has been introduced in [34]. The system has been designed to play the role of a guide in a historical site and is based on a virtual character, named "Carletto" who is a spider having an anthropomorphic aspect. It uses the DramaTour methodology for information presentation that is based on the assumption that a character acting in first person and sharing with the user both the location and the time of the visit can provide a strong effect of physical and emotional presence. The content is presented to the visitor in a location-aware manner and the elements

of the presentation are labelled with metadata associated to both their content and their communicative function.

A robotic system capable of interacting with human users in order to inform them about issues regarding the cultural heritage domain in an engaging and amusing manner has been illustrated in [35]. The system is capable of interacting in natural language with the user to explore a knowledge base both in a traditional, rule-based manner and also making sub-symbolic associations between the user queries and the concepts stored in the Knowledge base. Two different and complementary AI approaches, namely sub-symbolic and symbolic, have been implemented and combined to reach these capabilities. Furthermore, when the required information is not available, the robot tries to find it on Internet.

### 4.2. Interaction Based on the Tactile Channel

The ARIANNA framework (pAth Recognition for Indoor Assisted NavigatioN with Augmented perception) uses the camera of the smartphone to follow a path painted or sticked on the floor, and, as a consequence, exploits the vibration signal to provide a feedback to the user. In this way, the smartphones are used as mediation instruments between the user and the environment. Such a solution allows a visually impaired or blind person to autonomously visit a museum, contributing, consequently, to social inclusion. Some special landmarks (e.g., QRcodes or iBeacons) along the path code can give additional information detectable by the camera [36].

The cultural heritage environment in several cases is enhanced with 3D models reproducing artifacts, that can be obtained also by using 3D printing devices. They enable a tactile experience allowing visitors, in particular for people with vision problems, to deepen the different details of an artwork. They can also be used to provide access to something that is inaccessible or to reconstruct something that does not exist anymore. Often the 3D model is enriched with audio to provide further information about the artwork [37,38].

### 4.3. Interaction Based on the Visual Channel

3D models can be digitally defined when the purpose is to extend the experience with augmented and virtual reality. Virtual and mixed (augmented) reality technologies [2,9] allow the user to enrich his/her fruition experience or visit sites that are not physically accessible [39]. The virtual immersion usually involves sight and has a minor impact on other senses. Music, sounds, narration and interactive speech could be elements easily included in the immersive user's experience and could give a practical, emotional impact. Tactile [40], smell [41] perceptions require more complex devices and technologies that are expensive and difficult to include in a consumer system. Generally, AR and VR systems that interact are better appreciated if they do not require user adaptation and only employ intuitive and natural human communication channels. Ideally, a multimodal experience could assure a compelling realism of an augmented or of a virtual environment and a realistic simulation of the user's presence in it. On the other hand, the user's multimodal perception has to take into account its affective impact [42]. For this reason, the monitoring, and understanding of the human emotional reaction [43] should be another aspect to consider to build dynamic and adaptable systems.

The potential offered by AR and VR systems can be significantly increased through the Internet of Things (IoT) [44], it is considered a key enabling technology for making environments smarter and more interactive [45–48]. Internet of Things (IoT) can enhance the fruition experience. Museum installations can be devised by using smart autonomous objects, that are objects augmented with sensing, processing, storing, and actuating capabilities and that can also exchange information with each other [49,50]. It is of interest to consider how AR can give an intuitive way to communicate with IoT objects. Main issues are related to three aspect regarding data management, viewer and display devices and interfaces with their interaction methods. The main problems related to data management consist of building a virtual objects dataset and giving information on the surrounding

environment for AR whereas, for IoT, there is a need to guarantee access to information available from sensors and a good quality of service. Viewer and display devices should avoid latency and registration errors for AR and less-intuitive context information for IoT. Finally, the interfaces are expected to give a fixed interaction method for AR and intuitively communication devices with a limited response time for IoT. However, the use of both AR and IoT results advantageous compared to a ubiquitous computer and a more straightforward AR approach. As a matter of fact, even if computers are always present in different forms, the user is forced to interact with different devices; differently, the augmented interaction allows using a unique computer connected to the systems to be controlled. Besides, by using both approaches, the user can interact intuitively for example, by means of a helmet with hand gestures; it results in a more natural interaction with surrounding world and allows interaction with devices even if they are not physically present or with objects that cannot be touched or handled, such as valuables or cultural goods. This approach also permits additional information to be proposed to the user depending on his/her preferences. To access the surrounding objects, the user receives the collected AR attributes by a server in which information are shared; then, the user by the AR browser program can connect all of the things exploiting prebuilt relationship mapping [44]. This approach in which things are available under a unified Web framework (also known as "webization of things") can be usefully exploited in cultural heritage framework with the development of AR devices, assuring the user a good experience.

In simple contexts such as objects control, either in situ or remotely, a graphical user interface (GUI) implemented on a smartphone could be appreciated, but it does not give the presentation of synthesized images directly to the human eye; differently head-worn display (of Head Mounted Display HMD) are expected to increase the users' attention [51]. These AR devices can be classified into optical and video see-through HMDs, depending on whether actual images are viewed directly by the user or via a video input. The video see-through head-mounted display (HMD) allows two image sources (the real world and the computer-generated world) by a dual-webcam module and an immersive HMD display whereas the optical see-through HMDs mix virtual objects, and the user can see through them (computer-generated world) [52]. These devices, together with sensors based on high-speed networking, can allow users to perceive natural AR experiences in the augmented reality environment with the communication method of AR datasets used for recognizing and tracking them anyplace and anytime [44]. Based on these considerations, the real environment can be added with virtual objects. The same approach is proposed by [53] adopting an RGB-D camera in conjunction with simultaneous localization and mapping (SLAM) to retrieve the user's movement trajectory and obtain information related to the real environment. The SLAM with monocular camera in AR context [54,55], where only some virtual objects are placed within a scene of the real world, removes the limitation of VR and avoids simulating every entity within the scene. In [53], segmentation and shape fitting of objects are used to construct virtual objects to replace (within the virtual world) actual objects encountered (within the real world). Even if the authors claim that this system is suitable only for single-room venues and applicable mainly to static environments, it seems interesting for increasing the experience of visiting museums because the presence of objects can be enriched with others related to the same theme or with multimedia content.

A more detailed study about the user interaction on Augmented Reality using Smart Glasses is proposed by [56] where the attention is focused on Epson Moverio [57] and Microsoft HoloLens [58]. It is observed that a narrow field of view (FOV) makes it difficult to visualize information of AR display; however, more intuitive interaction is achieved, including hand-gesture-based interaction. The see-through display is classified in (a) stereo rendering and direct augmentation, (b) non-stereo rendering and indirect augmentation with video background, (c) non-stereo rendering and indirect augmentation without video background. The first method superimposes virtual information directly onto its physical artifact through the see-through display of smart glasses; it requires a registration process. The second overlays virtual information indirectly onto the physical artifact in the image

taken by the smart glasses' AR camera, whereas the former presents only the virtual information on the display. The hand-gesture interaction is given by a depth sensor attached to the smart AR glasses; even if it requires a calibration, in the context of cultural heritage the method (a) can be exploited to show as example an image before restoration and then by gesture the user can rotate it to observe different perspectives. For this purpose, HoloLens reveal more suitable due to the wider FOV. It should be also underlined that the accuracy the hand gesture and visual registration in wearable AR is expected to be further improved in the next future [56] .

The role of ICT in tourism has been recognized as able to provide unique opportunities for innovative organizations to redesign tourism products to address individual needs and to satisfy consumer wants [59], its impact is leading to e-Tourism. In [59] is remarked that tourists have become more demanding, requesting high-quality products and value for their money and, perhaps more importantly, value for time; as a consequence, a corresponding answer is expected by new technologies. This approach is used in [60] for artefacts in the context of cultural heritage and is applied to Port wine in [61] showing that the success of using both VR and multisensory technologies to conceptualize solutions in which users can virtually experience a set of senses so that they are totally immersed in a more exciting and realistic experience. Recently, the paper [62] proposed a Mobile Augmented Reality (MAR) application containing historical information related to Ovid's life. This approach is aimed to give a different perception of enriched reality with a computer-generated layer containing visual, audio and tactile information in a virtual representation of a classical museum. The user is immersed in the scene by Real-time tracking and localization and 3D animations of Ovid created a feeling of interaction. A suitable sound reconstruction enriches the experience. Both AR and VR allow for reconstructing tangible and intangible cultural heritage [63]. The modelling of virtual humans is at the basis of museum exhibitions described in [63] and implemented in the context of European projects. Such systems require a motion capture system for acquiring body and facial animations. Animated characters can be embedded in virtual restitution of a historical site, allowing for a simulation of the social dimension. They can exploited to deepen a historical figure, a better understanding of his life and actions [63], giving back behavioral aspects such as posture, clothes, body movements and gestures. As in the case of the virtual Ada Lovelace [63], projected onto a big wall screen inside a tunnel and properly reacting to the presence and movements of people. The behavior and the intention of people (such as approaching Ada, going away or staying in front of her) are captured by a Kinect device, then processed and analyzed by means of Markov chains and statistical analysis. The system also integrates an emotion recognition system based on body movement analysis and a dimensional emotions model. The Ada reactions rely on a behaviour management unit.

*4.4. Interaction Based on the Artistic Channel*

The ADA intelligent room [64,65] is an artificial entity embedded in a multimodal, immersive and interactive space developed for the Swiss national exhibition; such a system is inspiring for our vision. The intention at the basis of ADA was to stimulate a debate about the implications of artificial brains in our society. ADA locates and identifies people by using vision, audition, and touch senses. It has an artificial skin made up of floor tiles made up of pressure sensors, neon tubes and a microcontroller. The system reacts to people interacting with sound effects obtained by a synthetic musical composition system, and with images, video, and games screened on a 360-degree ring of LCD projectors. ADA acts according to the processing of neural networks and expresses a sort of internal status and emotional tone of the space by using a ring of ambient lights, gazer lights with pan, tilt and zoom capabilities making up her "eyes". Other examples of artistic installations reacting to the presence and activities of people are Dune [66] and Pulse Room [67]. Dune [66] is a light landscape where sensors and microphones capture participants' footsteps and sounds: the installation expresses a sort of mood, when is alone it is sleeping, while in

the presence of people reacts with lights and sounds. Pulse Room [67] is an interactive installation composed of a hundred of clear incandescent light bulbs, uniformly distributed over the exhibition room, reacting to the heart rate of participants.

## 5. User Understanding

A cultural heritage guiding system should provide the most appropriate content according to the specific profile, even coarse, of the user to improve the cultural heritage visit's effectiveness or the information provided on a cultural domain.

Many data are usually available in this domain and both the background and the emotions felt by the user can improve the fruition of the user experience. Therefore, it is essential to model both the user and the particular context in which the user is located to provide most effectively the information that best fits the user's expectations. Furthermore, detecting the emotions that people feel by visiting a cultural heritage site or while they are watching an artwork is one of the most relevant issues to consider for making the experience of the user exciting, engaging, and effective.

In the two following subsections, we describe some of the works that contain relevant aspects in these two areas of cultural heritage. The goal is to give an overview of some of the works in the literature that somehow take into account, exploit or underline the relevance of the user understanding in cultural heritage applications.

### 5.1. User's Profiling

User profiling plays a crucial role in cultural heritage guides. In most approaches oriented to cultural heritage, various information can constitute the user profile, and this information can be exploited to make the experience more personalized and effective. The different uses rely on the specific application. It could be the analysis of explicit or implicit user inputs, the exploration and mining of contents in social media posts, the user's current position, different heuristics, the user's rating of items or topics, and so on. Some of the applications reported in the literature in the last years, that somehow exploit a user profile, are briefly recapped in the following.

Hayashi et al. [68] presented a system that collects images from the Internet to create a customizable "Personal Virtual Museum" system. The idea is to give the user the impression of exploring a museum as real, accomplished by simply drawing on images listed in a web browser. The user provides the image files and the metadata associated with them. Depending on the user's input, the system automatically creates a 3D environment and also generates captions. The system provides many ways for the user to provide images; one of the most common is the Wikimedia Commons.

Different ICT solutions designed and implemented in the framework of a tourism-oriented project called O.R.C.HE.S.T.R.A., aimed at delivering contents belonging to the cultural heritage of the historical center of Naples, have been illustrated in [69]. The system is relevant for our four pillars approach, since it suggest to adopt user profiling methodologies and a recommendation system, based on a Collaborative Filtering approach, both for individual users and for groups of tourists, able to filter possible resources and facilitate the decision-making process of users, inferring preferences through the analysis of data from "Facebook.com". This choice made it possible to include an automatic profiling of the user. This approach avoids boring the user with filling out questionnaires, automatically learning his preferences. Finally, a recommendation system helps the user to make his choices identifying the different Points Of Interest (POIs) that could constitute a personalized touristic route.

The authors of [69] extract user preferences from the content posted on social networks to obtain a recommendation related to a particular domain. The system collects data from the profile and evaluates the similarity between the current user and other users of the application. Then, considering only the most similar users, their preferences are exploited to generate predictions and suggestions for the active user. Moreover, each analyzed data

is associated with a time weight to decrease the relevance of older data, increasing the influence of more recent user behaviors.

The relevance of using social media data for user understanding has been also highlighted in [70], where a system exploiting a soft sensor-based approach [71] has been illustrated for detecting what the interest of the public is potentially focused on. The methodology has been presented to better address the resources a museum can have to promote an event or to detect which artwork needs more attention from the public or is commonly underestimated by people. In particular, the system takes advantage of soft sensors that use lexical semantic approaches for inferring sentiment and emotions expressed in posts of social media. The system focuses on Facebook public pages of museums and measures the sentiment (*positive* or *negative*), the emotions (*anger, disgust, fear, joy, sadness, and surprise*), and the strength of both a post and its comments, given by the number of "likes" and the number of shares, to give a synthetic measurement on the overall reaction of people to a given initiative or a given event.

A Big Data infrastructure for the management of objects related to the "Cultural Heritage" field was illustrated in [72]. With regards to the knowledge of users, two aspects are considered—the user profile, which is a static type of information, and its behavior in terms of "data logs". The static information is given by a tuple that considers ontological attributes and metadata that characterize the user profile. The User Data Log is a set of tuples associated with a set of data (e.g., timestamp) and attributes that may be useful in describing the behavior of the user or his/her activities concerning an object of interest.

A project concerning a technological infrastructure for the improvement of cultural resources, named *ArkaeVision* has been presented in [73]. The system is user-centered and it can give different manners of providing cultural heritage assets, including virtual representations of museums. The system makes it possible to adapt what is going to be presented to the user according to his/her profile. The profile is initially created through the completion of a questionnaire during registration to the platform. Subsequently, the profile is updated by keeping track of: (a) the choices made by the user during the visit, (b) the additional information that was explored, and (c) the time spent on the entire visit. In the end, the user can review his profile and he/she can participate in some games and challenges, according to the gamification paradigm. This allows the user's status to be continuously updated, adapting the users experience to their interests.

A customizable system in the Cultural Heritage domain is illustrated in [74]. The system is focused on the capability of tailoring the contents to be presented according to the cognitive characteristics of a user. The system consists of a repository and four main components: the "initialization" component, used to initialize the context, the "rules" module, that is used to generate a set of rules for personalization, the "user modeling" element, which is responsible for generating implicitly and transparently in real time, the user model, and finally the "personalization" component, which provides the personalized content to the user. In particular, the user modeling is based on classifiers that correlate the cognitive characteristics of the end user with visual interaction and behavior patterns; the system captures user interactions (e.g., mouse click, key pressed, blink, location, timestamp etc.) and collects data regarding gaze and saccades calculating appropriate metrics that can be used to model the user behavior in predefined classes, called "cognitive styles".

A hybrid recommender system in the cultural heritage domain is described in [75]. The approach is centered on User Profiling and it considers the activities of the user and his/her friends in social media as well as semantic knowledge provided by Linked Open Data (LOD). The approach extracts information from Facebook by analyzing the generated content of users and their friends, performs disambiguation by exploiting LODs, profiles the user as a social graph, and provides suggestions on cultural heritage sites or items in the vicinity of the user's current location. As for the User Modeling, a graph is generated whose vertices can be: "user", "place", "place coordinates", "category of the place"; while the edges can be "knows", "visited", "located in", "has category". In the case where social information is not available, the system nevertheless allows the user to define a profile

explicitly through a questionnaire by clicking on a series of images related to 37 different categories of POIs and then updated implicitly through feedback every time the user leaves a place. In this case, the user profile is given by a vector of weights whose values express the user's interests in a given category of POIs.

A methodology for improving the user experience in Cultural Heritage Institutions by capturing information from the user's social activities, rather than explicitly from user input, is illustrated in [76]. The CURE (Cultural User peRsonas Experience) methodology is based on three phases: "*data acquisition*", "*reasoning*", and "*dissemination*". The "data acquisition" collects data from various sources (mobile devices, social media, open access web data) and applies data mining and natural language processing techniques to obtain the user profile. Moreover, the "data acquisition" phase integrates data from different sensors and context-aware algorithms. During the reasoning process, the data are analyzed to create a user model, while the dissemination phase deals with spreading the profile in the application to provide personalized content for the user.

A methodology for adaptive and context-aware system design has been presented in [77]. The approach exploits a formal semantic model of intelligent environments and a behavioral approach based on rules. The system has been implemented in a real museum and collects different environmental parameters of the rooms like humidity, temperature, and position sensors. Furthermore, in our user understanding view, the approach is relevant also because it acquires knowledge about the user preferences explicitly for different characteristics like his preferred historical period, the preferred style, and other kinds of information like the user's predisposition to walk. Moreover, an implicit knowledge about the user is inferred by analyzing his browsing history and social network contents that have been published or reported. The environment is modeled by using a context ontology, while the visit path is characterized by a set of metadata and a set of context information. The system uses an importance ranking method as a recommendation strategy.

The relevance of understanding users is also highlighted in [78], where a complete and comprehensive survey on the so called "*personalized access to cultural heritage*" is provided. It gives an overview of the different user models (overlay, feature-based, content-based, list of items) that can also be dynamically adapted with various techniques, like heuristic inference, activation/inhibition networks, collaborative filtering, content-based filtering, and semantic reasoning. Moreover, in that paper, different techniques are also highlighted to match the user preferences with the specific kind of content to provide at the right time. Among these techniques, the survey reports condition-action rules, ranking, semantic reasoning, activation/inhibition networks, collaborative filtering, content-based filtering, social recommendation methodologies, models based on vector spaces, and hybrid approaches.

Finally, a set of challenges for user modeling and personalization are outlined in [79]; in particular, the scrutable user modeling and adaptation, which becomes very hard when the adaptation is based on collective behavior: data mining, machine learning techniques, and deep neural networks are not still capable of explaining the conclusions that lead to a specific recommendation. A second issue is the repeatability of the research results in different contexts. The third challenge is the meta-adaption capable of avoiding filter bubbles, that is, when the user cannot easily access items that are not recommended.

*5.2. Emotion Analysis*

As has been highlighted by [80], cultural heritage sites are places where visitors go not only to enrich their knowledge, but also to experience emotions. The research reported in [80] has shown that focusing the visiting to only recreational or learning goals is too simplistic: museums and heritage sites are exploited and visited in many various manners and with different objectives, depending on the specific purposes of each visitor. Furthermore, interpreting and understanding how people make use, interact and catch the essential elements of a cultural heritage site makes is important.

According to this premise, an automatic system for emotion detection and recognition involved in Human Computer Interaction (HCI) and Affective Computing (AC) can play

a crucial role [81]. These systems can catch the visitors' emotions and, subsequently, the information can be exploited to improve the visiting experience of the users.

Emotions characterize one of the most fundamental ingredients of human behavior. This aspect can be effectively exploited in human-computer interaction by using an artificial intelligent system at the basis of an advanced cultural heritage system, constituting an essential feature of one of our four pillars approach. The system could recognize the emotions of the users and exhibit an empathetic response emulating a truly intelligent behavior.

Human beings manifest their emotions in many forms, including both explicit, indirect and non-verbal means—facial expression, speech, using specific utterances; sometimes, the use of irony or sarcasm can make it very difficult to recognize the exact expressed emotion. According to a study illustrated in [82] four aspects are tackled in recognizing emotions—facial expression recognition, physiological signals recognition, speech signals variation, and text semantics.

To highlight the prominent role played by emotions in cultural heritage, a methodology aimed at researching the relationship between cultural heritage spaces, emotions, and visitors has been presented in [83]. A case study, based on "Tredegar House" in the United Kingdom, has been illustrated. It aims to determine whether the Tredegar House can be a site capable of eliciting affective experiences and, if so, of what kind. To answer this question, visitor physiological data is captured, a post-visit survey is conducted, and some ethnographic notes are also considered. Each participant at the site entrance wears a device that measures different parameters, such as blood volume, pulse, heart rate, electrodermal, temperature, and accelerometer. Cardiac and electrodermal activities were considered in the study. After crossing two gardens, visitors decide whether to continue the view as "lady" or "maid," following two different routes. At the end of the visit, participants filled out a questionnaire asking what visitors found interesting during their visit, the emotions they felt, both with open-ended and closed-ended questions, and were asked to describe how these emotions had been triggered. The goal of the study was to identify if there are any relationships between how an experience is subjectively reported and physiological response, highlighting the value of qualitative data combined with digital tools.

A mobile application that supports the visit of cultural sites associated with the person of the musician Giuseppe Verdi has been presented in [84]. The system exposes a human-machine interaction mode based on the user's mood while exploring various sites of interest. The content that is shown to the user is driven by the state, the mood, of the person using the proposed system. In particular, the app, called "VersoVerdi", visually identifies the user's mood, which can be "happy" or "unhappy". Based on this information, the system provides suggestions of content to display corresponding to the mood identified. The system can also suggest links between museums: each museum is metaphorically shown as a planet. The planets that make up the universe can be explored by moving the device in the air. Each planet provides directions to other related sites. The mood is determined through a system of image processing that can detect whether the user is smiling or not. The adopted solution uses only local resources embedded in a smartphone, without any remote server's processing resources. Depending on the mood, specific contents are unlocked, and new paths of exploration are suggested. A happy user will be enabled to explore all the positive content related to the museum he is interested in, but not the negative ones that are, therefore, hidden.

One of the problems faced in [69] concerns the fruition in additional information on the object that the tourist is observing without, however, diverting the user's attention. To this end, an emotional eye-tracker, called E.Y.E.C.U., has been designed and implemented. It detects the details of what is observed by the visitor and, in real-time, shows any additional information associated with the work of art. An interesting aspect is that the system requires neither the wearing of special devices, nor the execution of particular gestures or unnatural poses. It is merely a remote gaze detector that determines head posture and tracks pupil position. The proposed approach is general and is based on the two-dimensional visual plane beyond the observed object to detect the portion on which

the user lingers his gaze. If the system detects such a condition, additional information is projected onto a region around the work on which the user has gazed. Moreover, pupil dilation is a reliable index in the analysis of emotional arousal; as a result, this parameter is also monitored to record the user's emotions during the visit. Once the visitor has moved away from the detection range of the camera, E.Y.E.C.U. gives a report indicating the sections observed and the related emotional reactions is generated.

Another work that is remarkable to cite is a first attempt to realize a mobile recommendation system in the cultural heritage domain, based on a facial emotion recognition approach, that exploits four emotions (happy, sad, surprise, and neutral) to estimate the user satisfaction of the content provided [85].

## 6. Gamification

Recently, there has been an increasing interest in the so-called gamification strategies. Gamification consists of applying game mechanics in non-game contexts where the game mechanics are blocks of interactions between a player and a game, for example, rules driving the player's actions and feedback (rewards, competitions). Several works show that gamification can have positive outcomes in terms of engagement and, in the specific context of learning, where this approach is even more exploited and acquired competencies and skills. In literature, gamification is also frequently used in the museum heritage context, creating a closer link between the visitor and the environment and supporting cultural heritage learning. For example, by playing the serious game MuseUs [86], the museum visitors can create their exposition. The game stimulates the attention of the visitors and the construction of personal narratives. Motion games have been proposed to explain and transmit intangible cultural heritage (ICH) resources and notions bound to bodily activities, such as serious games allowing users to learn and practice traditional Greek dance movements, where the user's movements are captured using a Kinect. They are subsequently compared to the dance of a professional dancer [87]. Obviously, the main target of gamification strategies is represented by young visitors, a significant part of people visiting cultural heritage sites. A situated tangible gamification installation aimed at enhancing student's cultural learning, fostering engagement and collaboration among them, is described in [88]. The installation is a replica of an ancient Egyptian tomb-chapel where it is possible to play three games to foster respectively architectural, historical, and artistic knowledge about the cultural heritage site; a light-emitting bar graph conveyed the progression of solving each of the games. In some cases, gamification is exploited to increase the motivation in labeling cultural heritage artifacts [89]. This tedious work, essential to describing ancient cultures' unique characteristics, becomes engaging thanks to the introduction of game components such as goals, levels and rewards.

Semantic Web and Internet of Things (IoT) are exploited in [50] to implement, in a real scenario, an engaging learning mechanism about a cultural heritage site. The approach relies on questions-answers games, where the multiple-choice questions are automatically created by extracting linked data from DBpedia and mining text descriptions about the exhibits.

In [90] is analyzed how the combination of different technologies and gamified experiences can motivate users to deepen their knowledge about museum sites. Among the cited works, ARCO (Augmented Representation of Cultural Objects) [91] is an architecture that allows museum curators to create virtual exhibitions and create quizzes. The museum visitors can access through a mobile application 3D models overlaid on the real world. Using the application can answer questions related to the AR or physical exhibits.

Teachers remarked increased motivation and engagement of their students using the tangible user interface game described in [92], combining manual activities with visual information in computer graphics, shown on a digital screen. I-Ulysses [93] is a virtual-reality game designed to support preservation and learning in cultural heritage. The work is aimed at educating players in the narrative and storytelling techniques of a classic literary work using an immersive virtual reality experience and proper game mechanics.

## 7. Our "Four Pillars" Perspective

In the previous sections, we outlined the leading technologies and methodologies proposed in the literature to enhance the user experience during the fruition of a cultural heritage site. We analyzed the state-of-the-art by grouping the examined solutions considering four main aspects that, in our opinion, are particularly important and of interest to our vision.

We conceive a cultural site as an intelligent entity if we extend it with elements belonging to a typical perception-reasoning-action cycle—sensors and smart devices, artificial intelligence-based methodologies for data analysis and innovative modalities of interaction. In the case of a museum environment or a site of cultural interest, human activities are the focus on the whole approach. Therefore, in our vision, an artificial entity's behavior, playing the role of a guardian/guide, should be driven by the desire of interacting with the museum visitors. It should provide them with pertinent information and establish both affective and amusing involvement to ensure an enthralling and exciting fruition experience.

Figure 1 provides a schema of our perspective, depicting our vision's four pillars as the main features (abilities) of the intelligent museum guardian. In this approach, the use of a Cognitive Architecture can help coordinating and managing the whole interaction between human users and the intelligent entity. The cognitive architecture should coordinate the four main aspects that we have highlighted, providing the intelligent guardian entity with basic components for knowledge representation, reasoning, and learning processes.
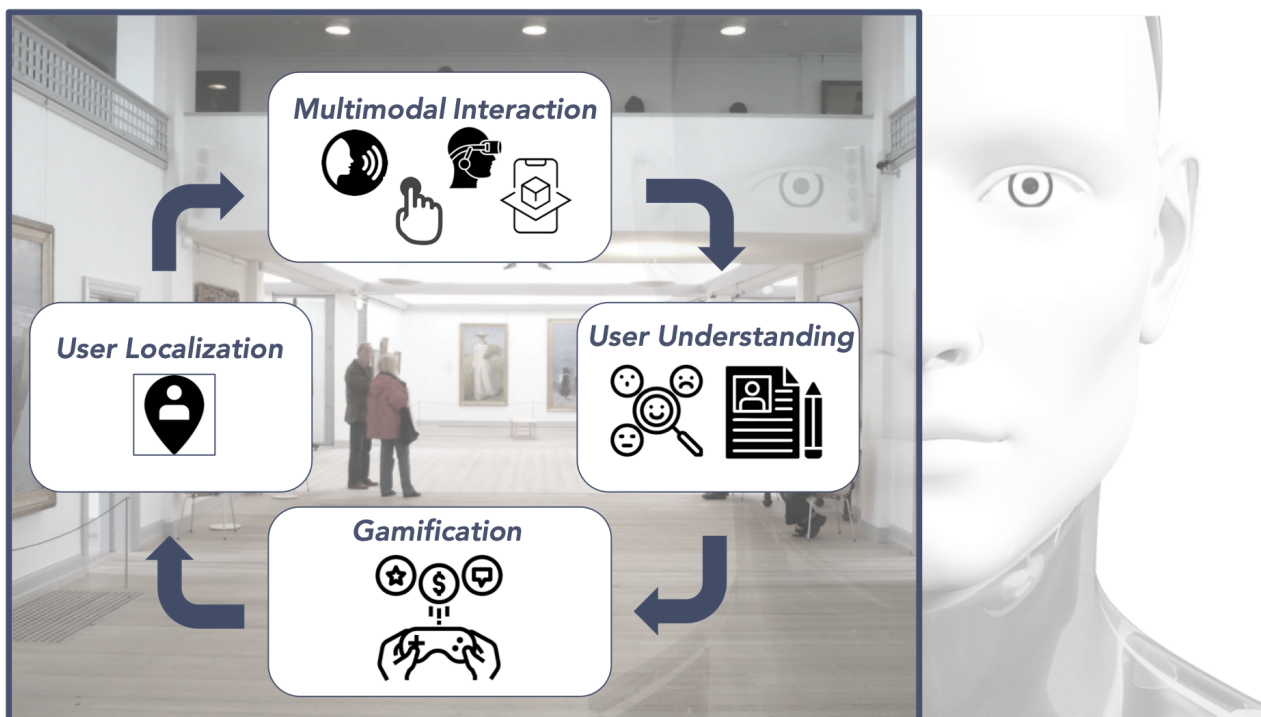


**Figure 1.** The four pillars of our vision.

Using a suitable cognitive architecture could allow the artificial guide to exploit socio-cognitive skills within a sort of Theory of Mind, (also known as mental perception, social common sense, folk psychology, social understanding). In fact, the intelligent guide needs the ability to recognize, understand and predict the human behavior regarding the underlying mental states such as beliefs, intents, desires, feeling, and so on [94,95].

The following features should be expected:

(a) To be responsive to people's expectations;
(b) To combine different models for knowledge representation and management to deal with different conceptualization and reasoning processes;

(c)    To make available natural and usable interfaces;
(d)    To exhibit a personality and to be conform to social rules;
(e)    To include a model of emotions;
(f)    To adopt engaging and creative strategies.

Table 1 maps the most inspiring (according to our vision) approaches discussed in the literature into the four blocks shown in Figure 1.

At present, we are investigating a cognitive framework named VESSEL, Valuable Environment for Social Supporting and Empathetic Linking, (see Figure 2).
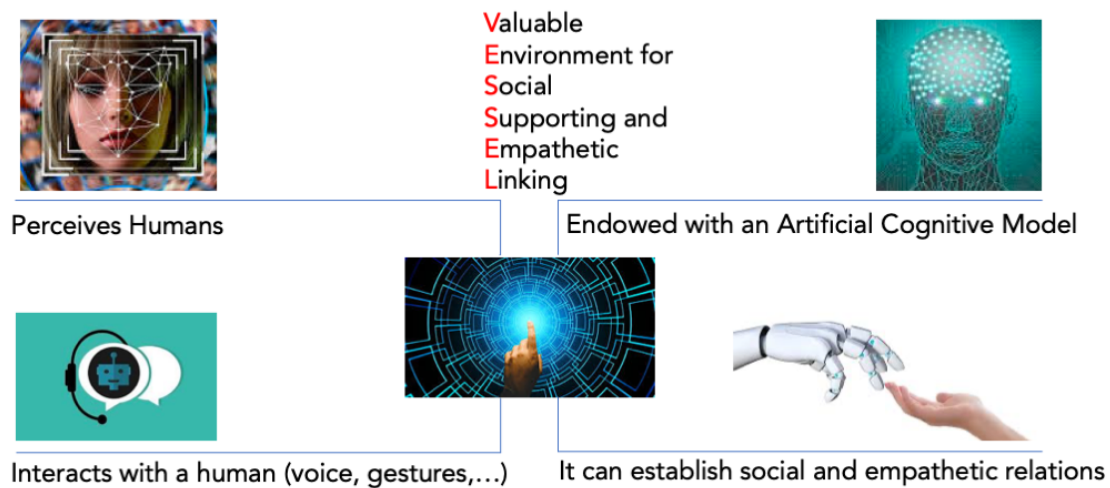


**Figure 2.** The Valuable Environment for Social Supporting and Empathetic Linking (VESSEL) framework.

VESSEL is conceived as a distributed architecture, with edge-computing modules (therefore localized on devices), external cloud modules, services, and processing modules on various computational resources (smartphones, tablets, servers). The framework is conceived to validate the agent's cognitive processes. Its implementation can be realized in a general robotic, or smart environment, platform. It is not necessary to implement it in an anthropomorphic robot. For example, in the context of cultural heritage, we imagine its main perceptive components and actuators of its body, as distributed in the environment, using different kinds of devices, sensors, virtual 3D elements, audio devices, and so on.

Some of the outlined aspects of our vision are being implemented in a project named VALUE (Visual Analysis for Localization and Understanding of Environments). In the setting of the project, visitors of a museum wear an AR (Augmented Reality) device, while a vision-based module (integrated with the visor) localizes the user in the environment by processing the camera's output embedded in the AR device, detects the gaze and identifies the object (or part of it) that the visitor observes (see Figure 3).
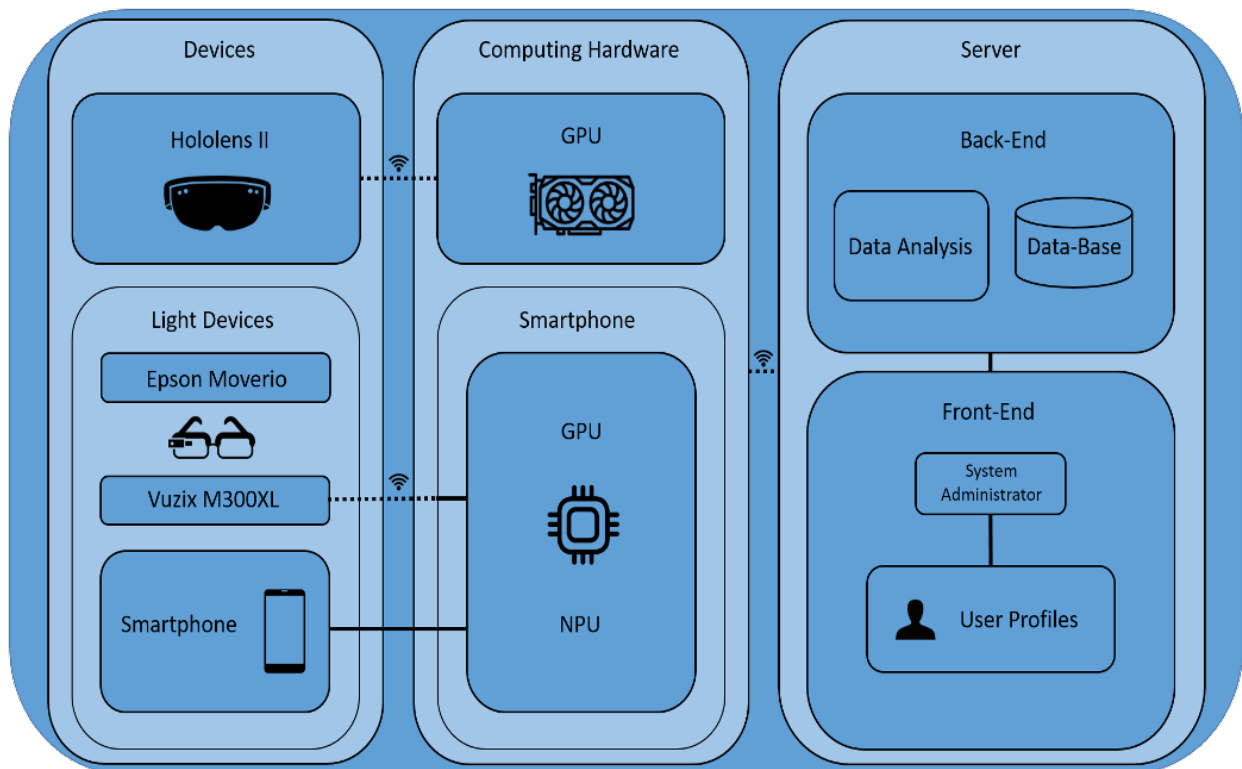
**Figure 3.** The architecture of VALUE (Visual Analysis for Localization and Understanding of Environments). A visitor wears an Augmented Reality (AR) device or smart-glasses equipped with software that allows the user to see and manipulate virtual objects and to talk with an avatar in natural language.

An avatar interacts with the user by providing information about an observed object, enriching the information with curiosities and stories. A natural language module supports the conversation, managing the dialogue.

The visitor can ask for further explanations and express his/her impressions. The system detects user emotions and sentiments from the dialogue, and, consequently, it adapts its behavior. Moreover, the AR device allows visitors to see virtual objects that enrich the museum's exhibition or view the actual items in different conditions (for example, before their restoration or the visualization of missing parts).

The setting up of a VESSEL-like system could take advantage of the experience made in the VALUE project. It could exploit the same technological setting, hence, having information about the user's point of observations (e.g., by analyzing his point of view and the gaze) and his reactions (e.g., by verbal expressions) to museum items.

The cognitive modules could then build on the fly a storytelling script, adapt it to the specific user desiderata and emotional state, and to adopt a gamification strategy to make the experience unique and engaging.

Naturally, the interactive system responsible for the cultural heritage site could also profile the visitor for commercial and merchandising purposes. By providing the system with cognitive modules that support self-explainability during verbal interaction, the user could be aware of the interactive system's behavior and aims.

**Table 1.** The literature with the most inspiring solutions with reference to the four pillars related to the authors' vision.

| | Localization | Multimodal Interaction | User Understanding | Gamification |
|---|:---:|:---:|:---:|:---:|
| Zafari et al [1] | X | | | |
| Yang et al. [15] | X | | | |
| Xiao et al. [16] | X | | | |
| Liu et al. [17] | X | | | |
| Kumar et al. [18] | X | | | |
| Povalac et al. [20] | X | | | |
| Scherhäufl et al. [21] | X | | | |
| Vasisht et al [22] | X | | | |
| Xiong et al. [23] | X | | | |
| Kotaru et al. [24] | X | | | |
| Baronti et al. [25] | X | | | |
| Gezici et al. [26] | X | | | |
| Kuo et al. [28] | X | | | |
| Huang et al. [30] | X | | | |
| Atzori et al. [31] | X | | | |
| Croce et al. [36] | X | X | | |
| Augello et al. [33] | X | X | | |
| Rossetti et al. [37] | | | X | |
| Samaroudi et al. [38] | | | X | |
| Sénécal et al. [63] | | | | |
| Delbrück et al. [65] | X | | X | |
| Hayashi et al. [68] | | X | X | |
| Barile et al. [69] | X | X | X | |
| Amato et al. [77] | X | | X | |
| Ardissono et al. [78] | | | X | |
| De Bra et al. [79] | | | X | |
| Smith [80] | | | X | |
| Saxena et al. [82] | | | X | |
| Hoare et al. [83] | | | X | |
| Roccetti et al. [84] | X | X | X | |
| Khellat-Kihel et al. [85] | | | X | |
| Su et al. [72] | | | X | |
| Bozzelli et al. [73] | | | X | |
| Raptis et al. [74] | | | X | |
| Sansonetti et al. [75] | | | X | |
| Konstantakis et al. [76] | | | X | |
| Coenen et al. [86] | | | | X |
| Grammatikopoulou et al. [87] | | | | X |
| Nofal et al. [88] | | | | X |
| Lee et al. [89] | | | | X |
| López-Martínez et al. [50] | | | | X |
| Cesaria et al. [92] | | | | X |
| Connor et al. [93] | | | | X |

## 8. Conclusions

The fruition and the accessibility of cultural heritage can be significantly improved, made more attractive to users through the use of new technologies and related methods based on localization and artificial intelligence, and in the meantime respecting the integrity of the cultural heritage assets.

The authors proposed an approach based on User Localization, Multimodal Interaction, User Understanding and Gamification, considered as pillars. Based on these, a review of existing literature has shown that many technologies and methodologies are already available on the market by smart devices; some others are expected to support this evolution soon. This confirms the practical, as well as theoretical, feasibility of our approach. It has been shown that, in order to provide any service to the user, the service itself must be located in the relevant context; then, the most appropriate content according to the

specific profile must be provided. This process goes through the user's profiling and emotion analysis. Based on this information, engagement can be improved. In particular, gamification gives positive outcomes, enhancing competencies and skills in the context of learning.

Instead of improving an existing context with technologies, this paper proposes to embed the fruition of cultural assets into the context of the four pillars described in this paper. Two experiments are underway to test the new approaches. The first one, named VESSEL, whose aim is to develop new artificial cognitive models and software modules to emulate artificial agents (virtual or embodied) embedded in a computational architecture. The second one is being implemented in the framework of a research project named VALUE. The visitor of a museum wears an Augmented Reality device; an avatar interacts with the user by telling facts and peculiarities of the object under observation, supporting a verbal interaction based on natural speech understanding. The visitor can ask for further explanations and express his/her impressions. Based on the dialogue, user emotions and sentiments are detected and, the system consequently adapts its behavior. The main scientific contribution of the paper consists of focusing the importance of gamification with use of Artificial Intelligence techniques to attract users.

## 9. Highlights

**Purpose of the paper:** *A perspective of cultural heritage fruition has been presented, and two experiences, currently in progress, are outlined, starting from an overview of emerging methods and technologies related to consumer electronics.*

**Methodology:** *The paper starts with a reasoned analysis of available methods and technologies applied to the cultural heritage context. They are mainly related to localization and user profile definition by IA-based approaches. We then identified the most important research databases on which we would perform our research (ACM, IEEE, Springer and Google Scholar). After a first selection and review of the collected works, we outlined those more meaningful for the four pillars. Finally, we summarized in a table the works most inspiring for our vision.*

**Findings:** *The research shows that technologies related to IoT are promising for performance, availability, and cost reduction. The user engagement is improved by gamification.*

**Research limits:** *Indoor localization is expected to be improved in granularity; sensors and systems related to AR as Smart Glasses should be used in addition to commonly available devices like smartphones.*

**Practical implications:** *The user interaction can significantly improve the fruition of cultural goods. Besides, the fruition details that cannot be seen in detail or touched can be encompassed.*

**Originality of the paper:** *We are combining different methods and technologies and proposing two experiences in progress as case studies.*

**Data Availability Statement:** Not Applicable, the study does not report any data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zafari, F.; Gkelias, A.; Leung, K.K. A survey of indoor localization systems and technologies. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2568–2599. [CrossRef]
2. Bekele, M.K.; Pierdicca, R.; Frontoni, E.; Malinverni, E.S.; Gain, J. A survey of augmented, virtual, and mixed reality for cultural heritage. *J. Comput. Cult. Herit. (JOCCH)* **2018**, *11*, 1–36. [CrossRef]
3. Steuer, J. Defining virtual reality: Dimensions determining telepresence. *J. Commun.* **1992**, *42*, 73–93. [CrossRef]
4. Nincarean, D.; Alia, M.B.; Halim, N.D.A.; Rahman, M.H.A. Mobile Augmented Reality: The potential for education. *Procedia-Soc. Behav. Sci.* **2013**, *103*, 657–664. [CrossRef]
5. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
6. Muñoz-Saavedra, L.; Miró-Amarante, L.; Domínguez-Morales, M. Augmented and virtual reality evolution and future tendency. *Appl. Sci.* **2020**, *10*, 322. [CrossRef]

7.  Yung, R.; Khoo-Lattimore, C. New realities: A systematic literature review on virtual reality and augmented reality in tourism research. *Curr. Issues Tour.* **2019**, *22*, 2056–2081. [CrossRef]
8.  Kimura, Y.; Manabe, S.; Ikeda, S.; Kimura, A.; Shibata, F. Can Transparent Virtual Objects Be Represented Realistically on OST-HMDs? In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 1327–1328.
9.  Bekele, M.K.; Champion, E. A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. *Front. Robot. AI* **2019**, *6*, 91. [CrossRef]
10. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine learning for cultural heritage: A survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [CrossRef]
11. Champion, E.; Rahaman, H. Survey of 3D digital heritage repositories and platforms. *Virtual Archaeol. Rev.* **2020**, *11*, 1–15. [CrossRef]
12. Bossavit, B.; Pina, A.; Sanchez-Gil, I.; Urtasun, A. Educational games to enhance museum visits for schools. *J. Educ. Technol. Soc.* **2018**, *21*, 171–186.
13. Paliokas, I.; Sylaiou, S. The use of serious games in museum visits and exhibitions: A systematic mapping study. In Proceedings of the 2016 8th International Conference on games and virtual worlds for serious applications (VS-GAMES), Barcelona, Spain, 7–9 September 2016; pp. 1–8.
14. Nikolakopoulou, V.; Koutsabasis, P. Methods and Practices for Assessing the User Experience of Interactive Systems for Cultural Heritage. Applying Innovative Technologies in Heritage Science. *IGI Glob.* **2020**, 171–208. [CrossRef]
15. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI: Indoor localization via channel response. *ACM Comput. Surv. (CSUR)* **2013**, *46*, 1–32. [CrossRef]
16. Xiao, J.; Wu, K.; Yi, Y.; Wang, L.; Ni, L.M. Pilot: Passive device-free indoor localization using channel state information. In Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems, Philadelphia, PA, USA, 8–11 July 2013 .
17. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 1067–1080. [CrossRef]
18. Kumar, S.; Gil, S.; Katabi, D.; Rus, D. Accurate indoor localization with zero start-up cost. In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, Maui, HI, USA, 7–11 September 2014.
19. Dargie, W.; Poellabauer, C. *Fundamentals of Wireless Sensor Networks: Theory and Practice*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
20. Povalač, A.; Šebesta, J. Phase of arrival ranging method for UHF RFID tags using instantaneous frequency measurement. In Proceedings of the 2010 Conference Proceedings ICECom, 20th International Conference on Applied Electromagnetics and Communications, Dubrovnik, Croatia, 20–23 September 2010.
21. Scherhäufl, M.; Pichler, M.; Müller, D.; Ziroff, A.; Stelzer, A. Phase-of-arrival-based localization of passive UHF RFID tags. In Proceedings of the 2013 IEEE MTT-S International Microwave Symposium Digest (MTT), Seattle, WA, USA, 2–7 June 2013.
22. Vasisht, D.; Kumar, S.; Katabi, D. Decimeter-level localization with a single WiFi access point. In Proceedings of the 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), Santa Clara, CA, USA, 16–18 March 2016.
23. Xiong, J.; Jamieson, K. Arraytrack: A fine-grained indoor location system. In Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), Lombard, IL, USA, 2–5 April 2013.
24. Kotaru, M.; Joshi, K.; Bharadia, D.; Katti, S. Spotfi: Decimeter level localization using wifi. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, London, UK, 17–21 August 2015.
25. Baronti, P.; Pillaia, P.; Chooka, V.W.C.; Chessabc, S.; Gottab, A.; Hu, Y.F. Wireless sensor networks: A survey on the state-of-the-art and the 802.15. 4 and ZigBee standards. *Comput. Commun.* **2007**, *30*, 1655–1695. [CrossRef]
26. Gezici, S.; Tian, Z.; Giannakis, G.B.; Kobayashi, H.; Molisch, A.F.; Poor, H.V.; Sahinoglu, Z. Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks. *IEEE Signal Process. Mag.* **2005**, *22*, 70–84. [CrossRef]
27. Oppermann, I.; Hämäläinen, M.; Iinatti, J. *UWB: Theory and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
28. Kuo, Y.-S.; Pannuto, P.; Hsiao, K.; Dutta, P. Luxapose: Indoor positioning with mobile phones and visible light. In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, Maui, HI, USA, 7–11 September 2014.
29. Armstrong, J.; Ysekercioglu, A.; Neild, A. Visible light positioning: A roadmap for international standardization. *IEEE Commun. Mag.* **2013**, *51*, 68–73. [CrossRef]
30. Huang, W.; Xiong, Y.; Li, X.; Lin, H. Swadloon: Direction finding and indoor localization using acoustic signal by shaking smartphones. *IEEE Trans. Mob. Comput.* **2014**, *14*, 2145–2157. [CrossRef]
31. Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Comput. Netw.* **2010**, *54*, 2787–2805. [CrossRef]
32. Chalmers, A. Experiencing the Multisensory Past. In *Mixed Reality and Gamification for Cultural Heritage*; Springer: Cham, Switzerland, 2017; pp. 359–370.
33. Augello, A.; Santangelo, A.; Sorce, S.; Pilato, G.; Gentile, A.; Genco, A.; Gaglio, S. Maga: A Mobile Archaeological Guide at Agrigento. Available online: http://hcilab.uniud.it/sigchi/doc/Virtuality06/Augello&al.pdf (accessed on 25 Feburary 2021).
34. Damiano, R.; Gena, C.; Lombardo, V.; Nunnari, F.; Pizzo, A. A stroll with Carletto: Adaptation in drama-based tours with virtual characters. *User Model. User-Adapt. Interact.* **2008**, *18* , 417–453. [CrossRef]

35. Augello, A.; Pilato, G.; Trecarichi, G.; Vassallo, G.; Vella, F. A Robotic Humanoid for Information Exploration on Cultural Heritage Domain. In *Rediscovering Heritage Through Technology*; Springer: Cham, Switzerland, 2020; pp. 47–65.

36. Croce, D.; Galioto, G.; Galioto, N.; Garlisi, D.; Giarré, L.; Inderst, F.; Tinnirello, I. Supporting Autonomous Navigation of Visually Impaired People for Experiencing Cultural Heritage. In *Rediscovering Heritage through Technology*; Springer: Cham, Switzerland, 2020; pp. 25–46.

37. Rossetti, V.; Furfari, F.; Leporini, B.; Pelagatti, S.; Quarta, A. Enabling Access to Cultural Heritage for the Visually Impaired: An Interactive 3D Model of a Cultural Site. *Procedia Comput. Sci.* **2018**, *130*, 383–391. [CrossRef]

38. Samaroudi, M.; Echavarria, K.R. Learning through experiences: Accessible fabricated dioramas for the visually impaired. In *International Conference on Immersive Learning, London, UK, 23–27 June 2019*; Springer: Cham, Switzerland, 2019; pp. 86–100.

39. Selmanović, E.; Rizvic, S.; Harvey, C.; Boskovic, D.; Hulusic, V.; Chahin, M.; Sljivo, S. Improving accessibility to intangible cultural heritage preservation using virtual reality. *J. Comput. Cult. Herit. (JOCCH)* **2020**, *13*, 1–19. [CrossRef]

40. Antlej, K.; Horan, B.; Mortimer, M.; Leen, R.; Allaman, M.; Vickers-Rich, P.; Rich, T. Mixed reality for museum experiences: A co-creative tactile-immersive virtual coloring serious game. In Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), San Francisco, CA, USA, 26–30 October 2018.

41. Harley, D.; Verni, A.; Willis, M.; Ng, A. Sensory VR: Smelling, touching, and eating virtual reality. In Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction, Stockholm, Sweden, 18–21 March 2018.

42. Riva, G.; Mantovani, F.; Capideville, C.S.; Preziosa, A.; Morganti, F. Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychol. Behav.* **2007**, *10*, 45–56. [CrossRef]

43. Marín-Morales, J.; Higuera-Trujillo, J.L.; Greco, A.; Guixeres, J.; Llinares, C.; Scilingo, E.P.; Alcañiz, M.; Valenza, G. Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. *Sci. Rep.* **2018**, *8*, 13657. [CrossRef]

44. Jo, D.; Kim, G.J. AR enabled IoT for a smart and interactive environment: A survey and future directions. *Sensors* **2019**, *19*, 4330. [CrossRef]

45. Jo, D.; Kim, G.J. Local context based recognition + Internet of Things: Complementary infrastructures for future generic mixed reality space. In Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST 2015), Beijing, China, 13–15 November 2015; p. 196.

46. White, G.; Cabrara, C.; Palade, A.; Clarke, S. Augmented reality in IoT. In Proceedings of the 8th International Workshop on Context-Aware and IoT Services, Hangzhou, China, 12–15 November 2018.

47. Phupattanasilp, P.; Tong, S. Augmented reality in the integrative Internet of Things (IoT): Application for precision farming. *Sustainability* **2019**, *11*, 2658. [CrossRef]

48. Gimenez, R.; Pous, M. Augmented reality as an enabling factor for the Internet of Things. In Proceedings of the W3CWorkshop: Augmented Reality on the Web, Barcelona, Spain, 15–16 June 2010.

49. Fortino, G.; Trunfio, P. (Eds.) *Internet of Things Based on Smart Objects: Technology, Middleware and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2014.

50. López-Martínez, A.; Carrera, Á.; Iglesias, C.A. Empowering Museum Experiences Applying Gamification Techniques Based on Linked Data and Smart Objects. *Appl. Sci.* **2020**, *10*, 5419. [CrossRef]

51. Kruij, E.; Swan, E.; Feiner, S. Perceptual issues in augmented reality revisited. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), Seoul, Korea, 13–16 October 2010.

52. Rolland, J.P.; Fuchs, H. Optical versus video see-through head-mounted displays in medical visualization. *Presence* **2000**, *9*, 287–309. [CrossRef]

53. Chen, Y.-S.; Lin, C.-Y. Virtual Object Replacement Based on Real Environments: Potential Application in Augmented Reality Systems. *Appl. Sci.* **2019**, *9*, 1797. [CrossRef]

54. Mur-Artal, R.; Montiel, J.; Tardos, J. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]

55. Engel, J.; Sturm, J.; Cremers, D. Semi-Dense Visual odometry for a monocular camera. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 1449–1456.

56. Kim, M.; Choi, S.H.; Park, K.Y.; Lee, J. User Interactions for Augmented Reality Smart Glasses: A comparative evaluation of visual contexts and interaction gestures. *Appl. Sci.* **2019**, *9*, 3171. [CrossRef]

57. Moverio BT-200. Available online: www.epson.com/MoverioBT200 (accessed on 5 April 2017).

58. MS HoloLens. Available online: https://www.microsoft.com/en-us/hololens (accessed on 20 March 2018).

59. Bethapudi, A. The role of ICT in tourism industry. *J. Appl. Econ. Bus.* **2013**, *1*, 67–79.

60. Di Giuseppantonio, P.; Matthews, J.L.; Matlock, T. Framing the past: How virtual experience affects bodily description of artefacts. *J. Cult. Herit.* **2016**, *17*, 179–187. [CrossRef]

61. Martins, J.; Gonçalves, R.; Branco, F.; Barbosa, L.; Melob, M.; Bessa, M. A multisensory virtual experience model for thematic tourism: A Port wine tourism application proposal. *J. Destin. Mark. Manag.* **2017**, *6*, 103–109. [CrossRef]

62. Boboc, R.G.; Duguleană, M.; Voinea, G.-D.; Postelnicu, C.-C.; Popovici, D.-M.; Carrozzino, M. Mobile augmented reality for cultural heritage: Following the footsteps of Ovid among different locations in Europe. *Sustainability* **2019**, *11*, 1167. [CrossRef]

63. Sénécal, S.; Cadi, N.; Arévalo, M.; Magnenat-Thalmann, N. Modelling Life Through Time: Cultural Heritage Case Studies. In *Mixed Reality and Gamification for Cultural Heritage*; Ioannides M., Magnenat-Thalmann N., Papagiannakis G., Eds.; Springer: Cham, Switzerland, 2017. [CrossRef]

64. Bullivant, L. Ada: The intelligent room. *Archit. Des.* **2005**, *75*, 86–90. [CrossRef]

65. Delbrück, T.; Eng, K.; Bäbler, A.; Bernardet, U.; Blanchard, M.; Briska, A.; Verschure, P.F. Ada: A Playful Interactive Space. Available online: http://ada.ini.ethz.ch/presskit/papers/ada-video-interact2003.pdf (accessed on 25 Feburary 2021).

66. Dune Installation. Available online: https://archello.com/project/interactive-landscape-of-light-dune-40 (accessed on 25 Feburary 2021).

67. Pulse Installation. Available online: https://www.lozano-hemmer.com/pulse_room.php (accessed on 25 Feburary 2021).

68. Hayashi, M.; Bachelder, S.; Nakajima, M. Automatic generation of personal virtual museum. In Proceedings of the 2016 International Conference on Cyberworlds (CW), Chongqing, China, 28–30 September 2016; pp. 219–222.

69. Barile, F.; Calandra, D.M.; Caso, A.; D'Auria, D.; Mauro, D.D.; Cutugno, F.; Rossi, S. ICT solutions for the OR.C.HE.STRA project: From personalized selection to enhanced fruition of cultural heritage data. In Proceedings of the 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, Morocco, 23–27 November 2014; pp. 501–507.

70. Pilato, G.; Maniscalco, U. Soft sensors for social sensing in cultural heritage. In Proceedings of the 2015 Digital Heritage, Granada, Spain, 28 September–2 October 2015.

71. D'Avanzo, E.; Pilato, G.; Lytras, M. Using Twitter Sentiment and Emotions Analysis of Google Trends for Decisions Making. Available online: https://www.emerald.com/insight/content/doi/10.1108/PROG-02-2016-0015/full/html (accessed on 25 Feburary 2021).

72. Su, X.; Sperlì, G.; Moscato, V.; Picariello, A.; Esposito, C.; Choi, C. An edge intelligence empowered recommender system enabling cultural heritage applications. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4266–4275. [CrossRef]

73. Bozzelli, G.; Raia, A.; Ricciardi, S.; Nino, M.D.; Barile, N.; Perrella, M.; Tramontano, M.; Pagano, A.; Palombini, A. An integrated VR/AR framework for user-centric interactive experience of cultural heritage: The ArkaeVision project. *Digit. Appl. Archaeol. Cult. Herit.* **2019**, *15*, E00124. [CrossRef]

74. Raptis, G.E.; Fidas, C.; Katsini, C.; Avouris, N. A cognition-centered personalization framework for cultural-heritage content. *User Model. User-Adapt. Interact.* **2019**, *29*, 9–65. [CrossRef]

75. Sansonetti, G.; Gasparetti, F.; Micarelli, A.; Cena, F.; Gena, C. Enhancing cultural recommendations through social and linked open data. *User Model. User-Adapt. Interact.* **2019**, *29*, 121–159. [CrossRef]

76. Konstantakis, M.; Michalakis, K.; Aliprantis, J.; Kalatha, E.; Moraitou, E.; Caridakis, G. A methodology for optimised cultural user personas experience-CURE architecture. In Proceedings of the 32nd International BCS Human Computer Interaction Conference, Las Vegas, NV, USA, 15–20 July 2018; Volume 32, pp. 1–8.

77. Amato, F.; Cozzolino, G.; Moscato, V.; Picariello, A.; Sperlí, G. Automatic personalization of visiting path based on users behaviour. In Proceedings of the 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), Taipei, Taiwan, 27–29 March 2017; pp. 692–697.

78. Ardissono, L.; Kuflik, T.; Petrelli, D. Personalization in cultural heritage: The road travelled and the one ahead. *User Model. User-Adapt. Interact.* **2012**, *22*, 73–99. [CrossRef]

79. De Bra, P. Challenges in user modeling and personalization. *IEEE Intell. Syst.* **2017**, *32*, 76–80. [CrossRef]

80. Smith, L. Visitor emotion, affect and registers of engagement at museums and heritage sites. *Conserv. Sci. Cult. Herit.* **2014**, *14*, 125–132.

81. Gaglio, S.; Infantino, I.; Pilato, G.; Rizzo, R.; Vella, F. *Vision and Emotional Flow in a Cognitive Architecture for Human-Machine Interaction*; IOS Press: Tepper Drive Clifton, VA, USA, 2011; pp. 112–117.

82. Saxena, A.; Khanna, A.; Gupta, D. Emotion recognition and detection methods: A comprehensive survey. *J. Artif. Intell. Syst.* **2020**, *2*, 53–79. [CrossRef]

83. Hoare, J. Heart on Your Sleeve? Emotion, Wearable Technology and Digital Heritage. In Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), San Francisco, CA, USA, 26–30 October 2018; pp. 1–4.

84. Roccetti, M.; Marfia, G.; Bertuccioli, C.; Marcomini, A.; Zanichelli, M.; Varni, A. Supporting cultural emotional browsing for museums: The versoverdi APP. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), San Jose, CA, USA, 15–19 July 2013; pp. 1–5.

85. Khellat-Kihel, S.; Pandolfo, L.; Pulina, L.; Grosso, E. On Leveraging Facial Emotion Recognition Techniques in Personalized Recommendations. In Proceedings of the Applications Of Intelligent Systems APPIS, Las Palmas de Gran Canaria, Spain, 8–12 January 2018; pp. 220–230.

86. Coenen, T.; Mostmans, L.; Naessens, K. MuseUs: Case study of a pervasive cultural heritage serious game. *J. Comput. Cult. Herit. (JOCCH)* **2013**, *6*, 1–19. [CrossRef]

87. Grammatikopoulou, A.; Laraba, S.; Sahbenderoglu, O.; Dimitropoulos, K.; Douka, S.; Grammalidis, N. An adaptive framework for the creation of exergames for intangible cultural heritage (ICH) education. *J. Comput. Educ.* **2019**, *6*, 417–450. [CrossRef]

88. Nofal, E.; Panagiotidou, G.; Reffat, R.M.; Hameeuw, H.; Boschloos, V.; Moere, A.V. Situated tangible gamification of heritage for supporting collaborative learning of young museum visitors. *J. Comput. Cult. Herit. (JOCCH)* **2020**, *13*, 1–24. [CrossRef]

89. Lee, J.; Yi, J.H.; Kim, S. Cultural Heritage Design Element Labeling System With Gamification. *IEEE Access* **2020**, *8*, 127700–127708. [CrossRef]

90. Bugeja, M.; Grech, E.M. Using Technology and Gamification as a Means of Enhancing Users' Experience at Cultural Heritage Sites. In *Rediscovering Heritage through Technology*; Springer: Cham, Switzerland, 2020; pp. 69–89.

91. White, M.; Mourkoussis, N.; Darcy, J.; Petridis, P.; Liarokapis, F.; Lister, P.; Walczak, K.; Wojciechowski, R.; Cellary, W.; Chmielewski, J.; et al. Arco-an architecture for digitization, management and presentation of virtual exhibitions. In Proceedings of the Computer Graphics International, Crete, Greece, 19 June 2004; pp. 622–625.

92. Cesaria, F.; Cucinelli, A.M.; De Prezzo, G.; Spada, I. Gamification in Cultural Heritage: A Tangible User Interface Game for Learning About Local Heritage. In *Digital Cultural Heritage*; Springer: Cham, Switzerland, 2020; pp. 411–422.

93. O'Connor, S.; Colreavy-Donnelly, S.; Dunwell, I. Fostering Engagement with Cultural Heritage Through Immersive VR and Gamification. In *Visual Computing for Cultural Heritage*; Springer: Cham, Switzerland, 2020; pp. 301–321.

94. Dautenhahn, K. Getting to know each other—Artificial social intelligence for autonomous robots. *Robot. Auton. Syst.* **1995**, *16*, 333–356. [CrossRef]

95. Infantino, I.; Augello, A.; Maniscalto, U.; Pilato, G.; Vella, F. A cognitive architecture for social robots. In Proceedings of the 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI), Palermo, Italy, 10–13 September 2018; pp. 1–5.