Mobile Augmented Reality for cultural dissemination

Raffello Brondi(1), Marcello Carrozzino(2), Franco Tecchia(3), Massimo Bergamasco(4)

(1) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, r.brondi@sssup.it
(2) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, m.carrozzino@sssup.it
(3) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, f.tecchia@sssup.it
(4) PERCRO, TeCIP Institute Scuola Superiore Sant'Anna, Pisa, Italy, m.bergamasco@sssup.it

Abstract—This paper presents the development of a Mobile Augmented Reality (MAR) framework, named XMAR, and of a prototype application, built on top of XMAR, and oriented to the cultural heritage field. The MAR framework has been developed to address the needs of outdoor markerless applications to be used in the field of cultural heritage. Furthermore, the framework had to meet the requirements of providing real-time registration using inertial sensors and GPS and of efficiently managing non-trivial tridimensional environments, transforming a typical smartphone in an augmented reality device. An application called LiTe has been designed and developed on top of the XMAR framework to provide an augmented view of Piazza dei Miracoli, one of the most famous artistic sites in Italy.

Mobile Augmented Reality; Markerless; Cultural Heritage;

I. INTRODUCTION

A huge amount of information about many cultural heritage sites has been extracted over the years by theoretic studies, though all these information is practically invisible to visitors. Augmented Reality (AR) systems can provide a solution to this lack, providing a valuable instrument for cultural heritage applications. They offer for example the possibility to see virtual reconstructions of ancient ruins in the same environment where the original buildings were. In addition, visual information can be combined with texts and multimedia information, providing interactive instruments and changing the learning process. The use of Mobile Augmented Reality (MAR) applications adds the ability to move within the sites while looking at the environment around you as it was when it was built.

Augmented Reality is a branch of the bigger research field of the Mixed Reality, which also includes Virtual Reality (VR), distinguished by the blending of virtual and real objects with the aim of augmenting the human perception of the surrounding environment and developing new metaphors of interaction. The sector of AR, as like as the one of VR, is closely related to the technological and technical progress. The goal of the first AR systems was the augmentation of an environment located in proximity of the user position. The technological barriers resulting from the devices used to implement such a system was the cause of that limit. Recently the improvements of mobile platforms, and in particular of smartphones, defined the new breed of the Mobile Augmented Reality (MAR) and opened a new thread in this research field. These devices include processors able to make relatively complex calculation and powerful graphic cards, inertial sensors and usually also GPS modules, the opportunity of fast connection via third-generation mobile telecommunications technologies and wireless connectivity in a unique small device. This set of characteristics offers a significant starting point on which it is possible to build MAR systems, which do not have mobility limits caused by the hardware used. Therefore, these systems can aim to provide the users with an augmentation of potentially unbounded environments, allowing him to move inside a large augmented environment.

The main motivation for the development of the presented framework has been the need for a tool for mobile devices capable of providing both a sensor based registration of the virtual environment, and able to handle and render non-trivial three-dimensional environments in real-time. The prototype application LiTe has been developed on top of this framework to verify the real capabilities of the framework. It is a MAR application for the "history browsing" of Piazza dei Miracoli in Pisa. The application resumes a previous research which has produced several multimedia information, included a tridimensional reconstruction of the artistic complex in several historical periods, from the 11st century up to now. The purpose of LiTe has been to provide the multimedia material produced by the previous research work through the use the Augmented Reality. The tridimensional models of the monuments can be seen overlaid on the video taken from the smartphone camera while the user is visiting the square. The user can then learn about the monuments around him through a new interaction paradigm characterized by a deep interactivity.

II. STATE OF THE ART

As part of the enhancement of cultural communication, there are many researches and applications aimed at exploiting, disseminating and making the artistic heritage easily accessible to non-specialist public. The use of new technologies is bringing about a substantial contribution to this field, and in particular the use of AR systems is increasingly proving to be valuable to this purpose.
Among the first researches in the field of Augmented Reality for Cultural Heritage it is noteworthy to mention ARCHEOGUIDE, a project aiming to provide a personalized electronic guide and tour assistant to cultural site visitors. It provides customized Augmented Reality tours and reconstructions of ruined cultural heritage sites in order to help visitors and scientists to better appreciate and enjoy the past glory of these sites. The system hardware is composed of a laptop and a head mounted display which communicate with a remote server containing the multimedia contents[6].

Another recent project with the aim of supplying a culture-oriented information system based on an AR system is iTACITUS. It combines itinerary planning, navigation and rich content on-site information, based upon a dispersed repository of historical and cultural resources. It enables a complete information system and media experience for historical interested travellers. It uses mobile computers to supply an augmented view of the cultural sites and spatial acoustic sounds to provide the user with an acoustic impression about how the place has been before[9]. The same team proposed an AR systems to be installed in museums to augment large wall-filling photographs of the real sites with interactive contextual annotations like 3D reconstructions, images and movies[8].

In the field of museums communication, it is historically important to mention "The Virtual Dig," made in 2003 by the Seattle Art Museum and the University of Washington Human Interface Technology laboratory. This interactive experience is based on an AR system for the presentation and exhibition of artifacts from Sichuan, and allows the visitors to perform some simple tasks, guided by a narrative, replicating the process of excavation and retrieval of archaeological finds. More recently, in 2008, the DNP Museum Lab has developed for the Louvre museum an AR guide [10], based on the Unifeye technology. The subsequent experience evaluation has shown a great public appreciation and the recognition of a substantially usefulness of the device. A subsequent evaluation experiment about the use of AR guides has been presented by CNAM [11].

The previous solutions use powerful and/or customized hardware to achieve their result, therefore they cannot be straight-forwardly implemented on devices with low capabilities such as smartphones. On the other hand, the system MARCH (Mobile Augmented Reality for Cultural Heritage) uses AR to assist visits to prehistoric caves using a smartphone [4]. The application is able to show a possible reconstruction of a painting on the cave walls but it relies on visual markers to perform the registration of the virtual painting with the real environment.

On the contrary, the system presented in this paper, does not make use of markers and can rely just on the hardware available on a commercial smartphone device to perform the registration of the virtual environment.

III. MOBILE AUGMENTED REALITY CHALLENGES

Any AR-based application presents a great deal of challenges that must be solved. There are two main common issues that must be addressed when developing an Augmented Reality system:

- real time interaction of the virtual environment with the user;
- registration of the virtual environment with the real world.

These issues are a direct consequence of the definition of AR as given by Azuma [1], and they have been studied ever since the advent of the first Augmented Reality systems. The impact of these problems on the final result is strictly related to the hardware configuration used and their incidence may vary for different hardware solutions.

Mobile devices, generally defined as handheld display, represents an excellent alternative to other kind of Augmented Reality displays, like wearable head-mounted displays, particularly because they are minimally intrusive, socially acceptable, readily available and highly mobile; moreover, they are usually cheaper and often present in everyday life. The typical use of such displays is to generate images at arm reach and video-through is the preferred paradigm. Integrated video cameras capture live video streams of the environment that are overlaid by graphical augmentations before being displayed. However, these devices, due to their intrinsic structural characteristics, suffer from a number disadvantages that have to be taken into account when developing an AR system:

- Power consumption: they have a limited autonomy.
- Computational power: they cannot run applications requiring heavy computations or being memory-intensive.
- Screen: their small screen results in a restricted FOV.
- Components quality: especially in consumer products, components, such as the camera or the sensors equipped on the mobile, might not suitable for AR.
- Hands-free: as opposed to head-mounted or projection based displays, handhelds force users to always have at least a busy hand.

IV. THE XMAR FRAMEWORK

The XMAR framework, so far available for the Android Operating System, has been designed and developed [16] to allow the easy and rapid development of Augmented Reality applications overlaying 3D co-located information to outdoor unprepared environments. This kind of scenario proposes several issues especially on mobile platforms.
Concerning the registration of virtual environments with the real world, there are many Computer Vision techniques which have shown excellent results in similar contexts, although they are computationally too expensive and to date not suitable for real time usage on mobiles. Thus, given the actual state of the art of smartphone technologies, XMAR bases the registration process on the combination of data given by the GPS and inertial sensors, providing a good compromise between precision of the registration and computational cost.

The real-time rendering of the augmented 3D information takes place through a library purposely developed on top of OpenGL ES 2.0, taking full advantage from the several structures supplied by this API.

A. Registration

The registration is the procedure through which an AR system brings the digital information to be correctly co-located with the view of the real environment. Without this, it is usually difficult to trick the human senses into believing that computer-generated virtual objects co-exist in the same physical space as the real world objects. To address this issue the registration process has to estimate both the position and the orientation of the device in order to place the virtual camera in the same reference system of the real environment.

To estimate the geographical position of the user, the framework uses the GPS sensors now almost ubiquitously available in the last generation of smartphones. The acquired geographic coordinates are defined in the WGS84 coordinate reference system, used as default reference system by almost all the GPS, which defines the user position using the angular values latitude and longitude. These values are converted to the UTM coordinate reference system, which provide distances in meters.

Each measurement contains a confidence value, indicating how accurate that particular measurement is. To reduce repositioning of the virtual camera due to inaccurate locations, an incoming value read from the GPS is used only:

- if it is significantly newer (more than a minute) than the current one, else
- if it is more accurate than the current one, else
- if it is not older neither less accurate than the current

otherwise it is discarded. Furthermore, the framework applies a low-pass filter to these values to reduce the noise of the measurements.

The orientation of the smartphone is obtained as a function of the readings of the magnetometer and accelerometer sensors applying the deterministic attitude estimation algorithm TRIAD [5]. The attitude matrix is then used to evaluate the view vector in the virtual environment. Like the GPS, magnetometer and accelerometer are not usually sufficiently precise and the signal coming from them is heavily noisy, affecting the evaluation of the virtual camera placement. To reduce the noise effect all the values read are filtered through a low pass filter.

The filtering processes used to reduce the errors of both GPS and inertial sensors make the registration more stable but of course introduce a certain delay in the update of virtual objects. However, since during the use of the application the user is supposed to make relatively slow movements, therefore a small delay is tolerable and is supposed not to be strongly noticeable.

B. Rendering

Although recent smartphones have processors and graphic cards much more powerful of the first MAR systems, the amount of graphic information that need to be handled and their quality has greatly increased. While the first systems limited the augmentation to text labels or trivial geometries, today the expectation for these kind of applications is much greater than before, due both to the increased availability of computing power and to our addiction to a more and more realistic virtual reality. On a smartphone application the management and rendering of an heavy tridimensional scene can still be a serious bottleneck, especially if it has to react to the user stimuli and being updated in real time.

The XMAR framework uses a library, called MVRLib, to manage and render the virtual environments. The library, written in C++ and representing a partial porting of the visual features of the XVR technology [15], interacts with the graphic hardware through OpenGL ES 2.0 and offers a compact object oriented interface to handle the complexity of a 3D environment. To avoid data redundancy, MVRLib uses four different static manager classes which respectively handle shapes, materials, textures and shaders. The virtual environment is hold in a Scene Graph which allows to structure the scene in a bounding volume hierarchy. On top of this data structure the library efficiently implements collision detection [13] and view frustum culling [12] algorithms to respectively provide a method for the selection of 3D objects through ray casting technique, and to boost rendering performance avoiding unnecessary visibility test in the graphic pipeline.

V. LiTe

LiTe is a prototype application developed on top of the XMAR architecture. The application is based on a previous work carried out in June 2001 when, in concurrence with the reopening of the leaning Tower, a large-scale multi-media project was launched on the Cathedral Square of Pisa, universally known as Piazza dei Miracoli.
One of the resulting products of the project was the creation of a web site (available online at the address http://piazza.opapisa.it/3D/) which contains a desktop Virtual Reality application enabling the exploration in space and time of the famous historical site. Based on the various surveys performed over the years on the Cathedral Square, the final 3D model was built and geo-referenced inside the urban system, depicting the whole site starting from the 11st century up to nowadays [3], identifying six different relevant historical periods virtually reconstructed.

LiTe, starting from this previous work, proposes a new form of interaction allowing to interactively access all the available multimedia material produced in the context of the described project. The main goals of the LiTe application are to:

- provide an augmented view of the site;
- allow the user to travel through different historical ages;
- allow the user to interactively request additional information about what he/she is currently seeing.

The application has been designed to be easily portable. It does not make any assumptions on the virtual environment it has to deal with. Once a georeferenced polygonal mesh has been supplied, it is able to provide an augmentation of the environment. For this reason it is possible to use the application in all those situations concerning outdoor monuments or archaeological sites. The artistic heritage is indeed full of sites, like Piazza dei Miracoli, whose communication can be greatly enhanced by the adoption of such instruments.

The application is structured on three layers. The deepest one displays the images captured by the device camera. The middle layer defines an OpenGL surface where the graphic library draws the virtual environment. The virtual objects are rendered overlaid on the video and are placed coherently with the real environment thanks to the registration process. The uppermost layer contains a set of widgets and buttons forming the GUI allowing the user to control the application.

A. User interface

When LiTe starts, it requests to the user to enter an internet address from where to download the tridimensional model to display. The application assumes that the specified data defines a 3D environment containing a georeferenced object to be used as an augmented content for the images acquired by the camera, using the position of the virtual camera as described in IV.A.

After loading the virtual environment, the application turns on the smartphone camera and shows the user the video streaming overlaid with the registered virtual environment.

Users can use the provided GUI to switch between the available historical periods (and therefore change the displayed 3D environment), to enable or disable the GPS positioning (a feature that can be useful when the noise of the acquired data is too high, resulting in a disturbing swing of the virtual camera), and to switch between the two implemented augmented modes, which will be detailed in the following section. The application also allows to use the touch screen to obtain information about the displayed virtual object. When a touch is retrieved, the system generates two screen coordinates and queries the XMAR framework to verify if the click has occurred on a region of the screen occupied by a virtual object. The framework performs a pick correlation on that coordinate and, if the test succeeds, it returns a label describing the touched object and, if available, further information registered for that element. The retrieved content is then displayed on the screen.
**B. Full VS semi augmented view**

The tests carried out in the real site show that the registration process still suffers from a lack of precision. In the worst cases, the virtual environment was too unstable due to the fluctuation of the acquired values, and the user immersion experience was damaged.

The precision of the registration process is strictly related to the hardware the smartphone is equipped with, and is also influenced by the surrounding environment. The GPS, for instance, may present precision issues due to the presence of buildings or other entities blocking radio signals from satellites.

To improve its usability, LiTe implements two different paradigms of navigation:

- fully augmented.
- semi augmented.

The first one represents the full augmentation of the square, where the tridimensional objects are shown overlaid on top of the video data coming from the camera. The latter, instead, does not render at all the virtual environment but is used to navigate through the real environment and query the application to obtain information about what is currently observed. The virtual environment indeed, even if not rendered, is still registered and when the user touches the screen to obtain information, the framework performs the pick correlation as previously described and displays the returned information as labels overlaid to the video stream. Thus, the semi augmented solution allows the user to receive only 2D augmented information about the surrounding environment; disabling the rendering of the 3D environment, however, although resulting in a somehow reduced experience, can be useful in case of a poor accuracy of the registration. The accuracy needed to obtain a good correlation between the touch zone and the virtual elements is in fact much less challenging, also thanks to the small dimension of the screen.

**VI. RESULTS AND FUTURE WORKS**

The application, tested on a Motorola Milestone A583 equipped with Android 2.1 Operating System, was generally very well accepted by test users, despite of noticeable precision issues. The GPS module equipped on the smartphone presented a poor accuracy and is heavily influenced by the environment. Position values coming from the GPS and inertial sensors are filtered by the architecture to reduce the effect of errors, but the results are not always satisfactory. These limitations are mainly related to the currently available low-price components and should be likely addressed by technology advancements in the immediate future. Smartphones are in fact conceived to be a product for a wide market and have to satisfy many different demands. Price and form factor considerations often lead to use cheaper or smaller components instead of more advanced solutions. As far as the 3D real-time rendering is concerned, the XMAR architecture has proved to be able to manage complex virtual environments. The frame rate obtained while rendering the tridimensional model of Piazza dei Miracoli, composed by about 50K triangles held in a scene graph of 561 different objects, is about 17fps, when using textures, and about 20fps, without textures.

Future work will focus on the improvement of the registration process. The latest smartphones are equipped, in addition to GPS and accelerometers, also with gyroscopes. These sensors allow the creation of more sophisticated filters, such as the Kalman filter[14], which should help to achieve better results in terms of precision and stability. Given the steady increase in computing power in recent smartphones, we will also consider and evaluate computer vision algorithms (such as SIFT tracking, PTAM, etc.) for the registration of virtual environment to test their applicability in this context.

**VII. ACKNOWLEDGMENTS**

The 3D Website of Piazza dei Miracoli was funded by Opera della Primaziale Pisana and Fondazione Cassa di Risparmio di Pisa, and realized together with VRMedia S.r.l, Andrea Brogi, Scuola Normale Superiore, Liberologico, Soprintendenza ai Beni Culturali di Pisa, Delcaldo WebStudio.

The research on Mobile Augmented Reality was partially carried on in the framework of the St@rt project, funded by Regione Toscana.

**REFERENCES**


