Teleimmersive Archaeology: Simulation and Cognitive Impact

Maurizio Forte[1], Gregorij Kurillo[2], and Teenie Matlock[3]

[1] School of Social Sciences, Humanities and Arts, University of California, Merced
[2] Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley
[3] School of Social Sciences, Humanities and Arts, University of California, Merced
mforte@ucmerced.edu, gregorij@eecs.berkeley.edu,
tmatlock@ucmerced.edu

Abstract. In this paper we present the framework for collaborative cyber-archaeology with support for teleimmersive communication which aims to provide more natural interaction and higher level of embodiment. Within the framework we create tools for exploration, interaction and communication of archaeologists in a shared virtual environment. Users at different geographical locations are captured by a set of stereo cameras to generate their real-time 3D avatars. The proposed framework is intended to serve as a virtual simulation environment where advanced behaviours, actions and new methodologies of research and training in archaeology, cognitive science and computer science could be tested.

Keywords: Cyberarchaeology, Teleimmersive Remote Collaboration, Shared Virtual Environments, Cognitive Impact.

1 Introduction

Teleimmersive Archaeology is a joint research project between the University of Berkeley (Teleimmersive Lab) and the University of California, Merced (Virtual Heritage Lab), supported by a CITRIS grant. The scope is the creation of virtual collaborative systems using teleimmersive technologies for real-time performances in the interpretation and reconstruction processes in archaeology. The methodological approach is based on the simulation process, in other terms on the idea that cyber-archaeology constitute the core of the interpretation capabilities in the digital simulation of the past (Forte, 2008).

One of the key issues in the collaborative-participatory activities is the role and the behaviours of all the actors involved in the process. Therefore factors such as the sense of presence, embodiment, gestures, interaction, motion capture, spatial sharing, 3D design and virtual tools influence deeply the level of learning and communication.

In the last decade several projects can be counted in this field: ARCHAVE (Acevedo et al., 2001); VITA: Visual Interaction Tool for Archaeology (Benko et al., 2004); SHAPE (Hall et al. (2001), LAVA, Laconia Acropolis Virtual Archaeology (Getchell et al. 2009). In addition other projects have explored the educational potentialities of the virtual communities such as Second Life (Nie, 2008). Further applications

have been focused on the 3D Web collaborative systems such as the case of the FIRB project (Forte and Pietroni, 2008), using Virtools DEV and Virtools Mutiuser Pack© by linking three different archaeological sites. In this case, all collaborative activity was online and with pre-determined 3D graphic libraries.

Although the massive multi-user environments seem appealing for such applications, they are currently unable to provide users with truly immersive experience or sufficient flexibility to construct the type of complex framework that we propose. The users of the former technologies were mainly observers of virtual replicas of ancient worlds, not active participants contributing to the reconstruction and interpretation processes.

2   Teleimmersion

Achieving an immersive experience in collaborative environments requires providing a visual experience similar to that delivered by reality. Traditional immersive virtual reality systems often use avatars, to represent human users inside the computer generated environments. Pre-modeled avatars, however, have several limitations with respect to body movement dynamics, gestures, eye contact and other subtle communication via body language and facial expressions. Likewise, the existing video conferencing technologies fail to properly preserve the eye gaze which has been shown to be an important factor for remote video-based communication (Fullwood and Doherty-Sneddon, 2006).

In our work we move further from the avatars and apply stereo reconstruction to capture 3D representation of users in real time to facilitate visual experience similar to reality (e.g. face-to-face meetings), where users are able to establish eye contact and use their body to communicate and interact (e.g. pointing at objects). The developed 3D reconstruction framework has been successfully used previously in remote dancing applications and learning of Tai Chi movements (Bailenson et al. 2008).
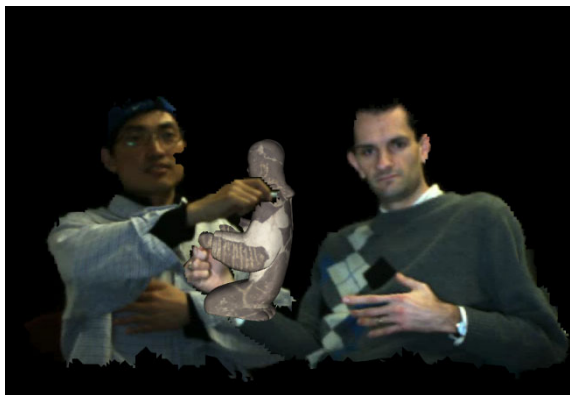

Fig. 1. Two users are interacting with a laser scanned statue in a shared virtual environment. Integrated 3D video provides a virtual position of each user, allowing users to point at different features of a 3D model as if they were sharing the same physical space.

3   Framework Overview

Our cyberarchaeology framework seamlessly combines computer vision and virtual reality (Fig. 1). The prototype application supports rendering and interaction with various 3D models, real-time interaction with different input devices, exchange of multi-media data streams for communication (i.e. audio, video and 3D video). The collaborative framework is built upon Vrui VR Toolkit, developed by Kreylos (2008) at University of California, Davis, which can run on a wide range of virtual reality hardware with support for different display and input device technologies.
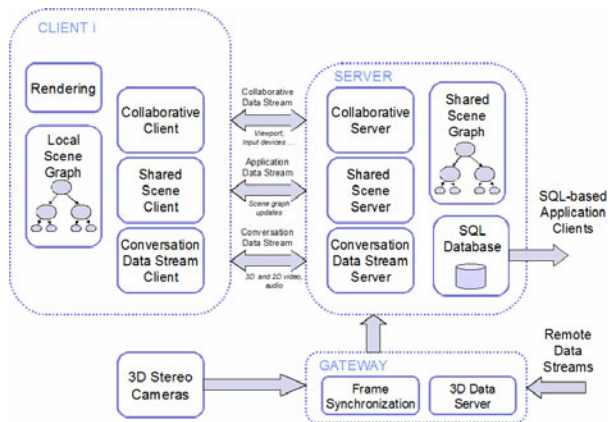
Fig. 2. A simplified block diagram of the 3D teleimmersive application for collaborative interaction in a shared virtual environment

3.1   Shared Scene Graph

The collaborative virtual environment is based on shared scene graph to describe the spatial relationship of objects and facilitate interaction between remote users. The scene graph is maintained on a central server which can connect to a spatial database. Whenever changes are made to the scene (e.g. adding/deleting objects, moving objects etc.), the clients receives the updates for their local scene graph representations. The scene graph also allows for efficient rendering. Current implementation with vertex buffer objects allows for display of 1 million triangles with the frame rate of 60 FPS (GeForce GTX 8800 graphics card).

The scene graph encodes different properties of the scene (e.g. geometry, texture, metadata) through a hierarchical scheme of inter-connected nodes of different types as follows:

Transformation Node defines position and orientation of its child nodes with respect to the other transformation nodes higher in the hierarchy. The transformation is described with six parameters. Nested nodes allow objects to be linked together.

Geometry Node describes object geometry through a list of vertices and a list of indices of the corresponding triangles. The data is used for building vertex buffer

objects (VBO) which allow for efficient rendering. The geometry node also encodes the object bounding box, which is used for the collision detection.

Texture and Material Nodes are comprised of object textures and material properties. Several high resolution texture file formats are supported in the application. The textures can also be dynamically switched, for example to alternate between the original and reconstructed surface of a wall painting. We implemented the surface material properties defined by the OpenGL standards.

Object Node is a group node that can incorporate several geometry and texture/material nodes defining a particular object. The clients currently support only OBJ/Wavefront 3D file format with the ability to use several different texture formats. The object node can be easily extended to support other 3D formats by modifying the file loading function.

Grid/Height Map Node is used to render surface grids for emphasizing different surfaces or creating a height map that defines the landscape of the archaeological dig. The map can be texture mapped with the images of the landscape to create more realistic rendering.

Metadata Nodes support rendering of images and text that can be attached to different artefacts in the virtual environment. The metadata can contain information on object geometry, location, short description, and images. Currently the metadata cannot be edited within the application.
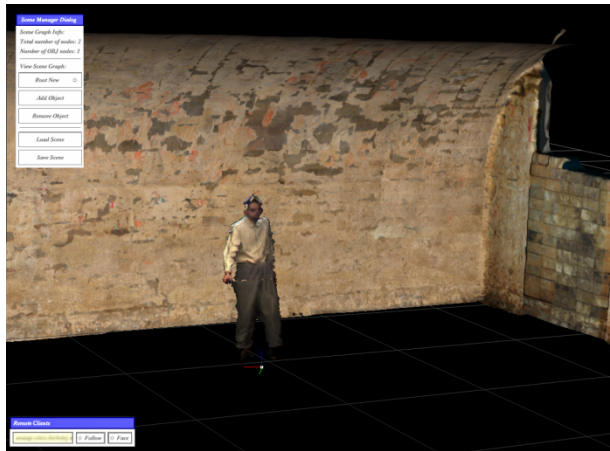

Fig. 3. Locally, users observe the virtual world in the first person perspective. Remote users are represented by their 3D avatars captured by one or more stereo cameras.

3.3   Navigation and Interaction

To explore the virtual environment, users navigate and interact with 3D models in the first person perspective. The presence of the remote users is accomplished through the rendering of their 3D avatar generated by the stereo cameras. The location of their avatar corresponds to the current location of the user's virtual view point.

At any time, individual users can switch to the other user's point of view or select face-to-face mode for direct conversation. The latter functionality will bring the local user in front of the remote user to facilitate a view similar to a video conferencing. Remote users can, however, work independently in the shared virtual environment. To prevent inconsistencies any two users cannot move the same object at the same time. A lock is placed on the node and its children if another user is already interacting with the object. The lock is assigned on first-come-first-serve basis.

The framework in connection with Vrui VR Toolkit features a wide selection of tools for interaction with the environment:

Navigation Tools provide a variety of ways to move around the virtual environment (e.g. flying, surface navigation).

Measurement Tool can be used to perform dimensional and angular measurements to capture the geometry of scanned objects. The measurement tools can also be used to measure spatial relationships between the objects. Fig. 4(a) shows the measurement of features on a small statue from a Western Han Chinese tomb.

Lighting Tool incorporates a virtual flashlight which can be used to relight parts of the 3D scene or point at salient features. The relighting can enhance underlying details of the scanned artifacts. Fig. 4(b) demonstrates the use of a virtual flashlight to enhance spatial details of a laser scanned model of a mask from Mayan city of Copan. In the future we will add ability to place static lights along the scene to more precisely control the illumination of the objects.

Annotation Tool allows users to draw 3D curves to mark different geometrical features and communicate them to the remote users. The annotation tool can also be used to quickly acquire a 2D or 3D sketch of patterns or objects. Fig. 4(c)(d) show an example of sketching the pattern on a scanned tile of a Western Han Chinese tomb and the corresponding geometry.

Dragging Tool is used to move objects in the 3D space. Object picking action is determined through a collision detection algorithm between the dragging tool selection ray and pre-calculated object bounding box. User can interact with an object only if another user has not already picked the same object. Movement of objects in different direction can be controlled independently through a dialog (when using a mouse) or through direct device interaction (when using 6 DOF input device). If the input device tracking and the stereo cameras are aligned, the hand of the avatar will be in contact with the object while it is being manipulated by the user.

Object Selector Tool is used to select objects and perform different actions related to the local functionality, such as changing object rendering style (e.g. texture, no texture, mesh only), retrieving object metadata, focusing current view to object principal planes etc. The selector tool allows selection of several objects simultaneously while different actions can be performed on selected objects.

3.4   3D Video Capture and Rendering

The avatars of users integrated with the virtual environment are created in real-time by the 3D stereo algorithm (Vasudevan et al., 2010). This algorithm performs accurate and efficient stereo computation by employing fast stereo matching through an adaptive meshing scheme. The output of the stereo reconstruction is a 3D mesh which is compressed and sent from each stereo camera to the local gateway. The achievable
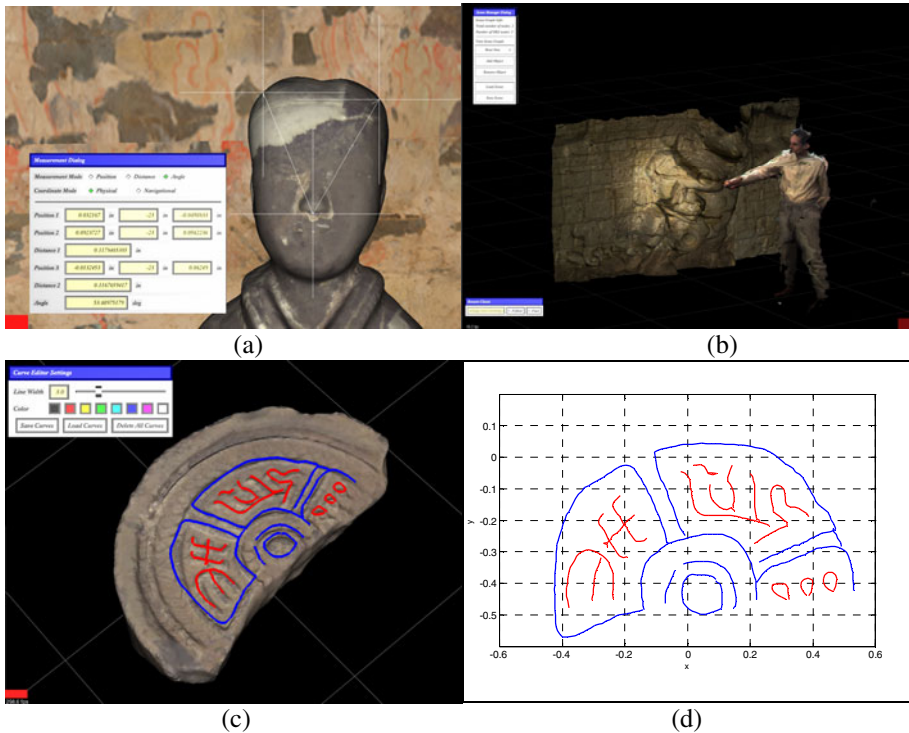
(a)   (b)

(c)   (d)

Fig. 4. (a) Dimensional and angular measurement can be performed on the virtual artefacts to capture their geometry. (b) Remote user is interacting with a virtual flashlight to enhance the underlying details of the laser scanned model of a mask from Mayan city of Copan. (c) Annotation tool is applied to mark important features and communicate them remotely to other collaborators. (d) Data from the annotations can be extracted in a form of a 3D sketch for subsequent analysis.

frame-rate is about 25 FPS on images of 320x240 pixels and about 12 FPS on images of 640x480 pixels. The accuracy of the reconstruction and is typically between 1 cm to 3 cm. To increase the fidelity of the reconstructed users, we also apply dynamic texture mapping. Several stereo views can be combined through calibration and blending to increase the workspace of the interaction and to provide 360-degree capturing.

3.5   Results

In this paper we present results from two different experimental setups (Fig. 5) connected over the internet. For the first setup, we used the teleimmersion platform at University of California, Berkeley (Vasudevan et al., 2010), which has several stereo clusters, each connected to a four core server, to perform 360-degree stereo reconstruction. The system is integrated with a tracking system (TrackIR by NaturalPoint) that tracks position and orientation of a Wii Remote (Nintendo). The Wii Remote is used as a 6 DOF input device for interacting with the virtual environment. The second

setup consisted of a single Bumblebee 2 stereo camera (Point Grey, Inc.) positioned above 65" LCD screen. Users were able to interact with the environment with a 3D mouse. Users can change the hardware platform by simply modifying a configuration file.

At this stage, we did some preliminary experiments with the 3D archaeological data coming from a monumental Chinese Western Han Tomb (beginning of the first millennium AD). We recorded and documented the tomb with 3D scanners in the summer 2008 at Xi'an, China. The tomb, now closed to the public (because of serious problems of conservation) after the excavation, is now accessible only virtually. For the teleimmersive system, the tomb was reconstructed by laser scanner data and by the integration of high res textures of mural painting and 3D models of funeral goods (recorded by laser scanning), then re-contextualized in their original positions (Fig. 3). Even the corridor and the three chambers of the tomb have been studied in the collaborative systems, especially the architectural elements, the organization of the space and the relation between iconography, the funeral chambers, and the 3D model. The use of lighting and measurement tools with the capacity to move, share and compare the objects in the cyber space, to add visual layers and outlines in the wall paintings have considerably increased the simulation factors and the faculties for data interpretation. The involvement of different interactors in the cyberspace yielded new perspectives in the dialectics of the interpretation process and its multivocality.

Fig. 5. Two different teleimmersive platforms for interaction with the virtual environment (below) and their corresponding view in the virtual environment

3.6 Cognitive Impact

Cognitive scientists are studying how users interact and manipulate objects in this system. They will track the use and development of the system with various methods of investigation. In the end, the results will be used to develop robust learning tools and to improve the design and use of the system. In one line of cognitive research, the eye movements of multiple users will be tracked as they discuss objects and manipulate objects. Of interest will be how language directs the attention of users. In another line of exploratory research on users, the utility of a pure first person view (no avatar) and a pseudo-first person view (with avatar) will be tested. Here trade-offs related to

allotment of attention, ease of use, and awareness are expected. For instance, because of limitations due to cognitive load, users may visually attend to objects less when they are using an avatar to represent themselves. This will of course have implications for how well users learn material in the environment. However, ease of communication may be better with an avatar because users can point and use other gestures for disambiguating the speech stream. In yet another line of research, we will investigate how and when users point at objects (and locations of objects), including the consistency of their pointing and the spatial characteristics of their pointing.

3.7 Learning Process

Teleimmersive environments afford many learning opportunities because they allow users to collaborate remotely on many different types of projects. One obvious issue to study is how easily users can learn new information presented in the environment, and how to simplify or optimize the learning process. Setting up a robust environment that enables shared learning will allow users to collaborate more efficiently and to collectively learn further new material more readily in the future (for excellent discussion of scaffolding in technological learning environments, see Pea, 2004).

For the first cognitive study, we will use the Chinese Western Han Tomb, which was mentioned above, as learning environment. We will begin with a simple study that offers users either a first person view (no avatar to represent the self) or a third person view (avatar to represent the self). In each of these conditions, we will run 10 participants; each will randomly be assigned to either a director role or to an observer role. After they enter the tomb and get used to the environment, directors will be told to reconstruct funeral objects while the observer watches. To accomplish this task, the director will use a set of tools provided in a separate window. In the first person condition, the director will move his or her avatar to the objects, pick them up, and assemble them. In the second person condition, the director will simply touch the objects, pick them up, and assemble them. In this case, no avatar will be visible (to anyone in the environment). To learn how to successfully put the objects together, the director will look at a diagram that shows how the objects go together but no written text (so there is no explicit order of actions).

The task will intentionally be difficult to engage and challenge the users, and to adequately test for differences in performance. Directors will have as long as needed to complete the task. The director will let the observer know when he or she is finished. At that point, the observer will have to put the object together. Various methods will be used to test the success of learning in the avatar versus no avatar instructional phase.

One prediction is that the first person condition, directors will take less time to construct objects because the director will have a direct embodied experience and not have to worry about how the avatar appears. In turn, the observer who worked with the director in this condition should also perform the task more quickly. We also predict there will be fewer errors with the first person condition because there will not be an avatar to distract the visual attention from the objects and how they are assembled. Yet another prediction is that directors, and in turn, observers, will remember details about the task better in the first person condition because more of their attention will have been allotted to the objects and strategies involved in putting them together. Together, these results will provide valuable information about the utility of the system for learning, and cognitive impact.

Follow-up studies will be developed from this initial study. In one, we will cross the avatar/no avatar condition with user. This will enable us to study efficiency of learning with no avatar under any circumstances versus learning an avatar that can be seen only by the director or only by the observer. We will also extend the initial study to situations with multiple users, and in some cases, multiple users constructing an object together. This will be important to developing a system that takes multiple views and mental models into account. We will also test novice versus expert users of the system to determine optimal modes of instruction given learning stages.

4   Conclusions

The system is in prototypal phase and needs significant work to improve the tools, user interfaces and rendering. This learning platform will teach users how to interpret, reconstruct and communicate archaeological datasets using all the information available in a virtual participatory form, for instance, photos, movies, maps, 3D models, spatial data and texts. In the future, we will study the utility of interactions with avatar/no avatar versus first person interaction. In later lines of research, the eye movements of multiple users will be tracked as they discuss objects and manipulate objects. Of interest will be how language directions attention. Finally we will investigate how and when users point at objects (and locations of objects), including the consistency of their pointing and the spatial characteristics of their pointing. We will also study how well users remember materials they have learned in the environments and how well this information is retained over time.

Finally, the study and analysis of a virtual reconstruction process in archaeology will help the virtual community to re-contextualize and reassemble spatial archaeological data sets, from the first draft version (data not yet interpreted) to the final communicative level. The research activity will involve a bottom-up approach, i.e., the analyses of the archaeological remains as they were found on site, and a top-down approach, i.e., the reconstruction/interpretation of the data by cultural comparisons (for example architectural features, artefacts, frescos, styles, materials, shapes, and others).

References

1. Acevedo, D., Vote, E., Laidlaw, D.H., Joukowsky, M.S.: Archaeological data visualization in VR: Analysis of lamp finds at the great temple of Petra, a case study. In: Proceedings of IEEE Visualization Conference, San Diego, CA, pp. 493–497 (2001)

2. Bailenson, J.N., Patel, K., Nielsen, A., Bajcsy, R., Jung, S., Kurillo, G.: The effect of interactivity on learning physical actions in virtual reality. Media Psychology 11, 354–376 (2008)
3. Benko, H., Ishak, E.W., Feiner, S.: Collaborative mixed reality visualization of an archaeological excavation. In: Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR 2004), Washington DC, pp. 132–140 (2004)
4. Forte, M., Pietroni, E.: Virtual reality web collaborative environments in archaeology. In: Proceedings of the 14th International Conference on Virtual Systems and Multimedia (VSMM 2008), Cyprus, pp. 74–78 (2008)
5. Forte, M.: Cyber-archaeology: an eco-approach to the virtual reconstruction of the past. In: Proceedings of International Symposium on Information and Communication Technologies in Cultural Heritage, Ioannina, Greece, pp. 91–106 (2008)
6. Fullwood, C., Doherty-Sneddon, G.: Effect of gazing at the camera during a video link on recall. Applied Ergonomics 37(2), 167–175 (2006)
7. Getchell, K., Miller, A., Allison, C., Sweetman, R.: Exploring the Second Life of a byzantine basilica. In: Petrovic, O. and Brand, A. (eds.), Serious Games on the Move, pp. 165–180. Springer Vienna (2009)
8. Hall, T., Ciolfi, L., Bannon, L.J., Fraser, M., Benford, S., Bowers, J., Greenhalgh, C., Hellström, S.O., Izadi, S., Schnädelbach, H., Flintham, M.: The visitor as virtual archaeologist: explorations in mixed reality technology to enhance educational and social interaction in the museum. In: Proceedings of Virtual Reality, Archaeology, and Cultural Heritage (VAST 2001), New York, pp. 91–96 (2001)
9. Kreylos, O.: Environment-independent VR development. In: Bebis, G., et al. (eds.) ISVC 2008, Part I. LNCS, vol. 5358, pp. 901–912. Springer, Heidelberg (2008)
10. Nie, M.: Exploring the past through the future: a case study of Second Life for archaeology education. In: Proceedings of 14th International Conference on Technology Supported Learning and Training, Berlin, Germany (2008)
11. Pea, R.D.: The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. Journal of the Learning Sciences, 423–451 (2004)
12. Vasudevan, R., Zhou, Z., Kurillo, G., Lobaton, E., Bajcsy, R., Nahrstedt, K.: Real-time stereo-vision system for 3D teleimmersive collaboration. In: Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2010), Singapore (2010)