MobileTI: A Portable Tele-Immersive System

Wanmin Wu, Raoul Rivas, Ahsan Arefin, Shu Shi
Renata M. Sheppard, Bach D. Bui, Klara Nahrstedt
Department of Computer Science
University of Illinois at Urbana-Champaign
201 N Goodwin Ave., Urbana, IL 61820, USA
{wwu23, trivas, marefin2, shushi2, rsheppard2, bachbui2, klara}@illinois.edu

ABSTRACT

We present *MobileTI*, a portable tele-immersive system that merges 3D video representations of users in real time to enable remote collaboration across geographical distances. With portability as a main goal, we address the challenges in the camera setup, time synchronization, video acquisition, and networking in the design and implementation of the system. Having been deployed in public performances, MobileTI proves to be effective, efficient, and user-friendly. Our experimental findings in terms of technical performance and user feedback are presented.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video; C.2.4 [**Distributed Systems**]: Distributed Applications

General Terms

Design, Measurement, Performance

Keywords

Tele-immersion, Portability

1. INTRODUCTION

The last decade has witnessed the rising of Tele-Immersive (TI) systems. By capturing full-body human motion in 3D and exchanging the video data over the Internet, these systems can render the 3D photorealistic representations of remote users into a joint virtual space at interactive rates. The technology is particularly useful for physical activities where full-body motion is intensively involved. Applications have already been demonstrated in a variety of areas, such as Tai-Chi [1] and collaborative dancing [9], where remote participants use TI systems to interact, collaborate, and learn in a shared cyberspace. Other emerging applications are wheelchair basketball training, medical consultation, military rehabilitation, and different types of gaming.

MM'09, October 19–24, 2009, Beijing, China.

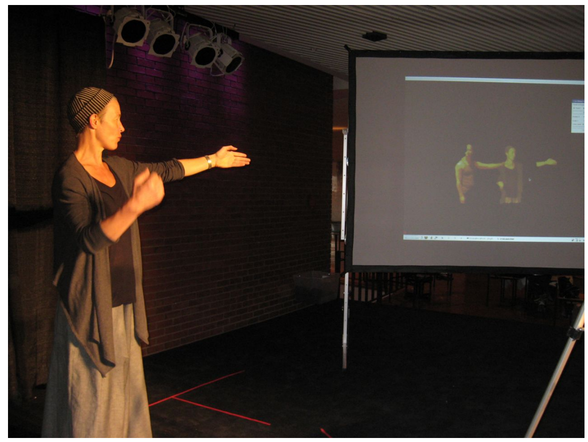


Figure 1: Public Performance using MobileTI

All of these applications demand high portability/mobility of the system, so that the users can easily take the equipment and set them up in the areas they deem fit. This turns out to be very challenging (if not impossible) with the existing TI systems. The existing systems create a fixed infrastructure in the controlled lab with static environmental settings on space, lighting, background,color and networking. The difficulty of maintaining those parameters in a public environment makes those systems less portable and difficult to move.

There are several barriers for portability on the current hardware and software design. The capturing setup is bulky and difficult to move with the handmade 3D camera clusters. Time synchronization of these cameras, required for stereo reconstruction, is achieved by connecting fixed hard wires from a server to those 2D cameras. Furthermore, the conventional 3D reconstruction algorithms are based on the point-cloud data representation, which has excessive bandwidth requirement, demanding large networking pipes such as Internet2.

This paper presents *MobileTI*, which embodies a suite of solutions to enhancing the portability of general TI systems. We address challenges in the key components of TI systems including camera setup, time synchronization, video acquisition, and networking. Specifically, we use compact 3D camera units on tripods to capture the scene, and a wireless channel for camera synchronization. We also present portability-oriented approaches in 3D video acquisition and networking among remote sites. To our best understanding, this is a first step toward building a portable and mobile TI system. We believe that without portability the existing TI systems can hardly gain wide application in reality.

The remainder of the paper is organized as follows. We first review the related work in Section 2, and then present the design and implementation of the MobileTI system in Section 3. We describe our experimental results in Section 4, and provide concluding remarks in Section 5.

2. RELATED WORK

Relatively few reports exist in the tele-immersion domain of real body and live art performance outside of the research lab.

Most video conferencing systems such as PolyCom, Microsoft Netmeeting and WebEx are easily portable and they aim to provide point-point or limited multipoint communication for desktop users. Only a single view is available with 2D web cameras. Some other systems like Virtual Auditorium [3] and Digital Amphitheatre [4] are designed for large scale video conferencing and provide telepresence experience beyond single 2D views. But these were designed primarily to use in a small office-desk environment rather than in an open public space such as a public auditorium or an indoor basketball court.

Coliseum [2], Virtue [8], Meta-Verse [10], the National Tele-Immersion Initiative [7], TEEVE [12] and UNC TI system [11] introduce 3D TI systems. The 3D view of a user is captured by multiple cameras, extracted from the background, then reconstructed and embedded into the virtual environment. Unlike other approaches, [11] and [12] produce multiple 3D video streams and transmit them over Internet2 yielding a much richer context in the reconstructed 3D cyber-space. Since the 3D representations of remote users are immersed into a virtual space, these systems allow interaction of body movement such as dancing [9], Tai-Chi learning [1] and basketball training. Hence, the system should be portable and easily installable in any public environment (e.g. a public auditorium or basketball court) for increased acceptance in those communities.

3. PORTABLE TELE-IMMERSIVE SYSTEM

3.1 Overview

Figure 2 gives an overview of the MobileTI system; for simplicity, only one pair of a sender and a receiver is shown. In practice, each participating site actually acts as both a sender and a receiver, thus the data exchange occurs both way. It is also straightforward to extend the architecture to multiple sites.

In the sender site, several 3D cameras are used for capturing. Their host PCs grab the 2D video frames and perform 3D reconstruction in real time, then send the 3D video data to a Rendezvous Points (RP) in the control center (a workstation). All cameras are time synchronized by a wireless trigger. The data is transferred from RP1 to RP2, then relayed to the renderers in the receiver site.

3.2 Physical Environment

TI systems have been widely studied in computer labs where the environmental conditions are controlled and where the lighting conditions are commonly fixed. Also as mentioned before, in previous TI systems the cameras are usually in a fixed location, surrounded by walls painted with

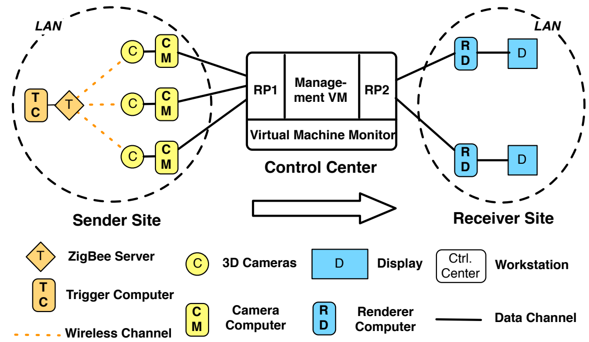


Figure 2: Overview of the MobileTI Architecture

distinctive colors and low sheen pigments. This is done with the purpose of reducing shadows and increasing the contrast between the foreground and the background planes of the images fed to the 3D reconstruction algorithms.

On the other hand, a portable TI system is required to support deployment in open areas where people, moving objects or variable lighting has a considerable effect in the quality of the 3D reconstruction of a video stream. To solve this problem, MobileTI uses a camera set up next to a screen where the the 3D video stream is projected. The field of view of the camera is marked on the floor so that the users are aware of this area. Immediately behind this marked area, a portable frame with a black curtain is erected. This frame is intended to reduce light changes, shadows and background movement that is nonexistent in a controlled environment (e.g. studio or lab). Lastly, the projector is set next to the curtain and out of the field of view of the camera. Figure 3 illustrates an example of the MobileTI setup.

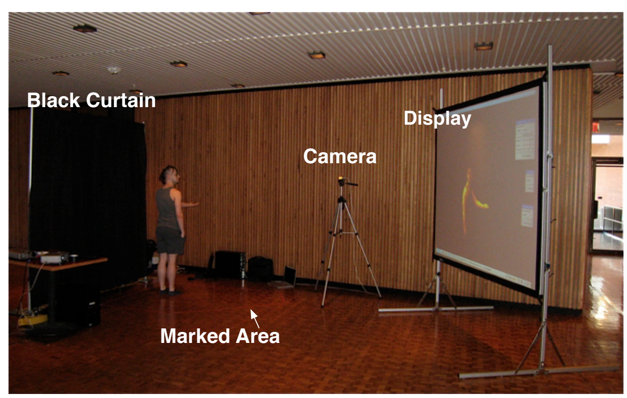
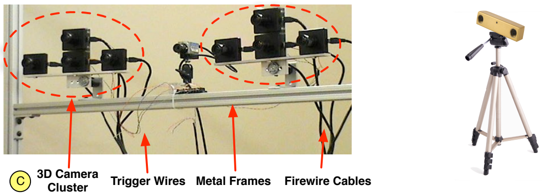


Figure 3: An Example Setup of MobileTI

3.3 3D Video Acquisition

3D video acquisition is comprised of two steps: (1) capturing and (2) 3D reconstruction. In existing implementations of TI systems, 3D video is captured by a group of four 2D cameras (Figure 4(a)). Images captured by the lower three cameras are used to calculate the pixel depth by stereo triangulation and the upper camera records the pixel color. The camera group needs to be calibrated and the relevant position must remain fixed after calibration (Figure 4(a)),



(a) Existing camera setup (courtesy of Dr. Peter Bajcsy)

(b) Portable 3D camera

Figure 4: Comparison of Camera Portability

which makes it only suitable for the static environment. Image frames captured by cameras are processed on a workstation for depth reconstruction. The high computation power and transmission bandwidth required by the point based 3D video also limit the mobility of the whole system.

MobileTI, in contrast, uses portable hardware and software components. On the hardware side, we use the stereo camera BumbleBee2 (Figure 4(b)) from Point Grey Research Inc. to replace the fragile camera group to capture the scene. There are several advantages of using commercial stereo cameras. First, the stereo camera has a smaller size but captures image frames at a larger resolution. Second, it is self-calibrated. There is no necessity to re-calibrate the camera in every new environment. For multiple cameras, we use Kurillo's fast method for external calibration [5]. Third, the stereo camera has the ASIC inside which accelerates the computation of pixel depth.

On the software side, a new mesh based 3D video representation is built based on the depth images generated by stereo cameras. The details of meshing algorithm can be found in details in [6]. The mesh representation not only fills potential holes in point clouds, but also successfully reduces the required transmission bandwidth to only a few Mbps. Thus, we can use a laptop to connect with the stereo camera for 3D video reconstruction and the wireless network for video streaming.

3D reconstruction also requires time-synchronized images from multiple cameras for the depth mapping to be correctly computed. To achieve synchronization, the existing TI systems use hardware triggering mechanisms that rely on coaxial cables to transmit a trigger signal from a computer to the I/O pins on each 2D camera (Figure 4(a)), which significantly degrades the portability of such systems. MobileTI uses wireless triggering so the wiring between the trigger server and cameras is avoided. In our implementation, we use the MicaZ motes with IEEE 802.15 protocols. In order to accommodate the different processing times on camera computers, triggering is designed to be adaptive. The MicaZ server waits for all cameras to acknowledge before firing the next trigger signal.

3.4 Networking

A portable TI system can be deployed anywhere, ranging from a newly designed smart building with wireless connection and Internet access to an old building with only power infrastructure. Therefore, a portable TI system must consider limited networking resources and the possibility of using private networking. Also ideally a portable TI system, by definition, needs to provide a good comprise between performance and the amount of hardware required.

Table 1: 3D Video Frame Rate (fps)

	Workstation with wired network	Laptop with wireless network
No user	20.2	18.6
1 user	19.7	17.3
2 users	18.3	15.3
More users	16.1	13.8

The MobileTI system is designed as a minimal deployment architecture in which each TI site connects directly to a centralized Gigabit Ethernet switch or Wireless Access Point. Additionally, to further simplify the system, MobileTI is set up in a private Local Area Network with private IP addressing. This increases the portability while simplifies the system configuration.

Also, to reduce the amount of physical hardware deployed, the TI system utilizes the concept of virtualization. Only one physical machine, the Control Center, running a Virtual Machine Monitor, hosts all the gateways required in the system to provide the Rendezvous Points (RP) for all mobile TI sites deployed. Figure 2 shows the architecture of the MobileTI system in which all the renderers and cameras connect to the virtualized Control Center. Virtualization also allows facilities for priority scheduling of real-time data streams as well as enhanced security options for private virtual spaces. The details of virtualization in TI systems are out of scope in this paper.

4. EXPERIMENTS

In this section, we describe our experimental findings on system performance (Section 4.1) and user experience (Section 4.2).

4.1 Performance

The technical performance of the system is evaluated from two aspects: frame rate and image quality. The frame rate is primarily determined by the number of pixels to be processed and the computation capability of the reconstruction server. Table 1 compares the frame rates generated with different equipments and scenarios. We find that the laptop with wireless communication is still able to provide the frame rate of 13.8 fps which is adequate for interactive communication.

Different from the frame rate, the image quality of the 3D video varies dramatically in different environments. Lighting conditions, background setups, etc, are all important factors that affect the final image quality on the screen. Although it is not easy to define and measure the image quality quantitatively, we summarize the following rules in how to improve the actual image quality of our experiments.

- Lights on the stage should be avoided causing heavy shadows on the background.
- Background selection should consider the avoidance of external light sources and open areas where people move frequently.
- The background color must be distinguishable from the color of clothes and skin of performers.
- The performers are encouraged to wear light color clothes.

4.2 User Experience

In order to validate the system from a user perspective, we asked two creative artists to design interactive, play-based activities for a public audience of about 60 people to engage with a remote partner through the tele-immersive system. The purpose of this experiment from a user experience perspective was to test the friendliness and interactivity of the system in a non lab based setting. We deployed the mobile TI system with two sites, 100 feet apart. The system was set up as part of a cultural event at the Krannert Center for the Performing Arts at the University of Illinois at Urbana Champaign.

As part of the activities designed, the audience, was asked to offer suggestions during a "simon says" type of game (one participant speaks instructions while the other physically follows the instruction) and to enter the tele-immersive space themselves to play the virtual game of"twister". The portable TI setup was extremely useful to both the users and the audience because it allowed for a comprehensive understanding of what tele-immersion offers in terms of bi-located physicality with a shared virtual environment. One of the users commented, "The audience wouldn't get it, wouldn't understand the point of what we were doing if they only saw the virtual space displayed on a screen and they wouldn't fully appreciate the real-time, interactive aspect if they didn't see the person moving in real life...it is this new kind of, eerie, dual performance of real and virtual and although they exist separately, they are not mutually exclusive."

With this experiment we observed that from a user perspective the innovative aspect lies primarily in the experience of moving with near immediate response from the system and the immediate reactions and therefore a shared experience of a person who is not physically there. Our experiment with the portable system provided an opportunity to see how an audience responds to the concept of tele-immersion. We observed that the tele-immersive environment is focused on the user experience and that the audience/third party experience is dependent on the presence and visibility of two components: the live participants and a screen displaying the shared virtual environment. This defines two different types of shared experience from the user perspective in which the objective dictates which setup is most appropriate. The performance appears to require a comprehensive setup in which the audience is capable of interacting in the physical space with the performers, while collaborative and communicative tasks such as Tai-Chi or physical therapy only require that the user himself/herself can see the screen.

In terms of friendliness we observed that initially people were not as comfortable in the system as in physical interaction, which was expected. It would take some time for first-time users to understand what the system can offer and what the differences are compared to face-to-face interactions. However, after a short learning process of a few minutes the users started to creatively utilize the system. We observed that after this learning period the users started to creatively engage remotely through various activities such as dancing and comic improvisation.

5. CONCLUSION

In conclusion, the design and implementation of MobileTI, a portable tele-immersive system is presented. We identify and address several challenges in different components of the existing TI system, and demonstrate the new MobileTI system by deploying it in public performance sites. We describe our experiences when building MobileTI and discuss the lessons learned. To the best of our knowledge, this work is the first attempt to enhance the portability of the existing TI systems. Our future plan is to conduct a more extensive evaluation of the system performance and user experience by deploying MobileTI in diverse open areas.

6. REFERENCES

[1] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S. H. Jung, and G. Kurillo. The effect of interactivity on learning physical actions in virtual reality. *Media Psychology*, 2008.

[2] H. Baker, N. Bhatti, D. Tanguay, I. Sobel, D. Gelb, M. Goss, W. Culbertson, and T. Malzbender. Understanding performance in Coliseum, an immersive videoconferencing system. *ACM Transactions on Multimedia Computing Communications and Applications*, (1), 2005.

[3] M. Chen. Design of a virtual auditorium. In *Proc. of 9th ACM Multimedia*, 2001.

[4] L. Gharai, C. Perkins, R. Riley, and A. Mankin. Large scale video conferencing: A digital amphitheater. In *Proc. of 8th ICDMS*, 2002.

[5] G. Kurillo, Z. Li, and R. Bajcsy. Framework for hierarchical calibration of multi-camera systems for teleimmersion. In *Proc. of 2nd ICIT*, 2009.

[6] G. Kurillo, R. Vasudevan, E. Lobaton, and R. Bajcsy. A framework for collaborative real-time 3d teleimmersion in a geographically distributed environment. In *Proc. of ISM*, 2008.

[7] D. E. Ott and K. Mayer-Patel. Coordinated multi-streaming for 3d tele-immersion. In *Proc. of 12th ACM Multimedia*, 2004.

[8] O. Schreer, N. Brandenburg, S. Askar, and E. Trucco. A virtual 3d video-conferencing system providing semiimmersive telepresence: A real-time solution in hardware and software. In *Proc. of International Conference on eWork and eBusiness*, 2001.

[9] R. Sheppard, M. Kamali, R. Rivas, M. Tamai, Z. Yang, W. Wu, and K. Nahrstedt. Advancing interactive collaborative mediums through tele-immersive dance (TED): A symbiotic creativity and design environment for art and computer science. In *Proc. of ACM Multimedia Interactive Arts Program*, 2008.

[10] S. Shi, L. Wang, K. Calvert, and J. Griffioen. A multi-path routing service for immersive environments. In *Workshop on Grids and Advanced Networks, in conjunction with CCGrid*, 2004.

[11] B. Wang, J. Kurose, P. Shenoy, and D. Towsley. Multimedia streaming via tcp: An analytic performance study. In *Proc. of 12th ACM Multimedia*, 2004.

[12] Z. Yang, Y. Cui, B. Yu, J. Liang, K. Nshrstedt, S. H. Jung, and R. Bajscy. TEEVE: The next generation architecture for tele-immersive environments. In *Proc. of 7th ISM*, 2005.