Large Area Robust Hybrid Tracking with Life-size Avatar in Mixed Reality Environment – for Cultural and Historical Installation

William Russell Pensyl, Daniel Keith Jernigan, Tran Cong Thien Qui, Hsin Pei Fang, Lee Shang Ping
Interaction & Entertainment Research Centre
Nanyang Technological University

wrpensyl@ntu.edu.sg

Abstract

We develop a system which enables us to track participant-observer accurately in a large area for the purpose of immersing them in a mixed reality environment. This system is robust even under uncompromising lighting condition. Accurate tracking of the observer's spatial and orientation point of view is achieved by using hybrid inertial sensor and computer vision techniques. We demonstrate our results by presenting life-size, animated human avatar sitting in a real chair, in a stable and low-jitter manner. The system installation allows the observers to freely walk around and navigate themselves in the environment, while being able to see the avatar from various angles. The project installation provides an exciting way for cultural and historical narrative to be presented vividly in the real present world.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual reality

Keywords: mixed reality, localization, hybrid vision and inertial sensor, cultural and historical installation

1 Introduction

Tracking black-and-white fiducial marker [Kato and Billinghurst 1999] has always been the conventional way in augmented reality (AR) or mixed reality of finding the six degree-of-freedom (6DOF) coordinates frame of the observer (or more precisely, the coordinates of the camera attached to him) with respect to the world coordinate, or the marker's coordinate. However, this technique suffers under uncompromising lighting condition; which causes jitters between frames (a big problem in AR), and even the complete loss of tracking altogether. Another problem with fiducial marker tracking is that tracking grows increasingly unstable as the view direction of the camera becomes perpendicular to the marker plane [Uematsu and Saito 2007]. In addition to this, for large area tracking, big marker or several markers would have to be dispersed in the environment, which can sometimes be visually unaesthetic.

Our work uses a hybrid approach – tracking an active marker while at the same time tracking the movement of the observer's (camera's) frame with the use of inertial sensor. The active marker is made up of an infrared (IR) light-emitting diode (LED) mounted on the user's head-mounted display (HMD). In our system, to detect IR LED, instead of using normal cameras, we are using Nintendo Wii Remotes as vision tracking devices. This low-cost device can detect IR sources at up to 100 Hz, which is very suitable for real-time interaction system. Furthermore, an inertial sensor which consists of 6DOF is attached to the HMD to detect rotations and movements of user' viewpoints.

Previous works such as [Sparacino et al. 1999] tracks humans in an environment "whose only requirements are good, constant lighting and an unmoving background". In [Nguyen et al. 2005], capturing the human body requires a green recording room with consistent lighting. Our system, on the other hand, works in most unexpected, varying, artificial, and/or ambient lighting condition.

2 Background

The motivation for developing this project is to allow people to experience pseudo-historical events impressed over a present day real world environment (we have set our project at the famous Long Bar at the Raffles Hotel in Singapore). It is an augmented reality multi-media art installation which involves re-enactment of famous people who frequented the bar in the early 20^{th} century. The piece will use augmented reality technology to develop both historical and legendary culturally significant events into fully interactive mixed reality experiences. Participants wearing head mounted display systems will witness virtual character versions of various notable figures, including Somerset Maugham, Joseph Conrad, and Jean Harlow, immersed within a real world environment modeled on the Raffles Hotel Long Bar they had once frequented. Through the application of research in tracking, occlusion, and by embedding large mesh animated characters, this installation demonstrates the results of the technical research and the conceptual development and presentation in the installation. Moreover, requiring that our work eventually be located in the Long Bar provides us with the motivation to create a system which can accommodate compromising lighting conditions and large open spaces.

[Song et al. 2004] developed a virtual reality (VR) system for digital heritage application. Their system allows user to navigate through a virtual heritage site, with the aid of a virtual tour guide. However, it does not track the user's whereabouts; the navigation is achieved by user pointing a device (brush) on an interactive screen. Our system differs in a few ways: it is an augmented and mixed reality system which users can really walk around within the area of installation while they are able to see virtual (but vividly realistic) human sitting in a real world environment.

This work prefigures new entertainment forms we refer to as interactive entertainment. In this new form, narratives, animation, film based presentation are crafted to occur in real world locations. The form is interactive and audience participative and moves away from the current passive entertainment forms in film, television, theatre and performance and place the viewer behind the "fourth wall" and immerses the viewer in an experiential way into the performance installation.

3 System Description

Figure 1 shows the system setup. The observer wears a HMD and can freely move around within the designated area. There is an empty chair, a table and other physical artifacts in the demo area. As the observer look at the chair, he sees a life-size, 3D animated human character sitting in the chair. The details of the system setup are described in the sections below.

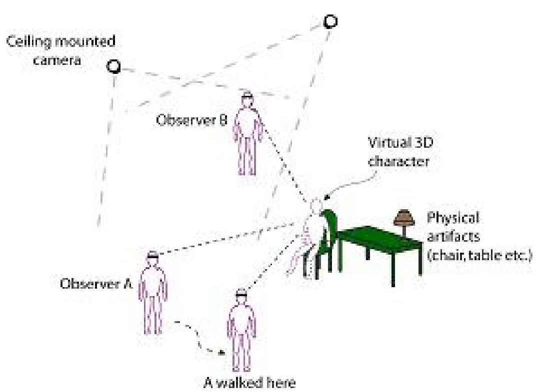

Figure 1. Setup of the large area robust hybrid tracking system. The chair, table and lamp are physical objects in the real world. The "human" in dashed-line and sitting in the chair is a 3D avatar.

3.1 Low-cost Vision-based tracking of head position

Conventional ARToolkit [Artoolkit] or MXRToolkit [MXRtoolkit] markers are not suitable for vision-based tracking in large area, uncompromising lighting environment. Jittering and loss of tracking due to the lighting condition seriously hampers accurate tracking, adversely impacting the audience's aesthetic experience. In view of this, we devise an active beacon by using IR LED as a position tracking device. Instead of employing black and white markers, the actors wear light-weight HMD which has an IR LED attached to it.

We use two Nintendo Wii Remote as vision tracking devices. The Wii consists of a monochrome camera with resolution of 128x96, with an IR-pass filter in front of it. The camera includes a built-in processor capable of tracking up to 4 moving objects (raw pixel data is not available to the host). 8x subpixel analysis is used to provide 1024x768 resolution for the tracked points [Wii Remote]. The Wii Remote cameras are installed in the ceiling to track the location of the actors (two cameras are sufficient to acquire the x, y, z of the HMD). This tracking provides positional information only and does not provide the orientation information of the head.

The advantages of using Wii Remote cameras are manifold: low-cost, easy setup, high "frame rate" (in fact only processed image – the coordinates of the tracked points are sent to the host) and wireless. The host computer is relieved from processing of raw image; and an optimal triangulation technique [Slabaugh et al. 2001] is all that is needed to obtain the depth information of the IR LED.

3.2 Inertial sensor-based tracking of head orientation

A small inertial sensor is installed underneath the IR LED. The sensor, which consists of accelerometers, gyroscopes and a digital compass, allows for the full roll, pitch and yaw tracking of the head orientation.

Apart from the 3DOF orientation readings from the inertial sensor, we integrate the acceleration of the accelerometer to obtain the position information (though corrupted with drifts); and then applied sophisticated sensor fusion algorithm (with Kalman filter) [Caarls et al. 2003; Hols et al. 2006] to combine with the vision-based position reading to obtain more accurate results.

3.3 3D Model, rendering and animation of the virtual human

In an AR application, it is crucial that the observer-participants become quickly engaged with the virtual content – otherwise they might lose their interest within seconds, adversely affecting their virtual experience. Therefore creating vivid, realistic 3D virtual human model in order to grasp participants' attention and keep them engaged for long period of time, has been one of the top priorities in this project.

Figure 2 shows the animation sequence of the virtual human. As can been seen, the model has got high quality textures and proportionate skeleton. The animation was key-frame based; and the body postures were obtained by optical motion capture method – by tracking optical beacons attached to a real human, thus making the motion look realistic. In addition to this, every single frame is generated by capturing the optical beacons, therefore essentially no computer-generated in-betweens are used, and all frames are key frames. This further enhances the smoothness of the animation.

Both the rendering and animation of the virtual human and object make intensive use of the Ksatria's kjAPI [Ksatria] game engine.

Figure 2. The animation sequence of the 3D virtual human. All are key-frame based and use no in-between. The motion is based on optical motion capture method.

3.4 System block diagram

Figure 3 shows the block diagram of the system. The peripheral hardware are shaded in blue.

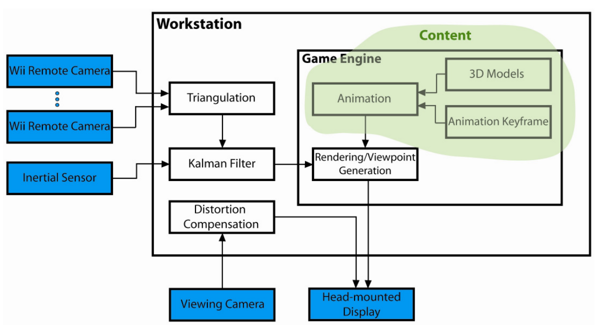

Figure 3. Block diagram of the system

We are using kjAPI game engine [Ksatria], developed by Ksatria Gameworks, to render visual contents and animations. 3D models and their animations are created in Maya and exported by plug-ins of the game engine. To communicate with Wii Remote through Bluetooth, Wii Yourself! library [Wii yourself] is used. In spite of using only one workstation, the system can operate at 30 frames per second and provide users an interactive and realistic entertainment.

4 Installation

The project is currently installed at the Art Gallery of Nanyang Technological University. The designated area is about 4m x 4m; it could be bigger or smaller depending on the camera lens and the strength of the IR LED. As shown in Figure 4 and Figure 5, the system works in the mix of studio lighting as well as varying outside lighting.

Figure 5 depicts the very interesting ingenious idea of using mixed and augmented reality technologies in cultural and historical installation – it breaks the time-and-space barrier between present day people and all things of the past.

Figure 4. The installation of the Long Bar. With naked eyes, observer-participants only see physical objects. The observer is wearing the HMD and looking at the table and chair.

Figure 5. View through the observer's HMD. It is as if the other participant is resting his hand on the shoulder of the virtual human. The lighting is dimmed as compare to that of Figure 4 but the tracking still works robustly.

5 Conclusions and future works

We developed a highly robust hybrid tracking system for use in augmented reality application. The system was designed in the application context of re-enactment of cultural and historical events and anecdotal stories of famous persons in the past. This technical cum artistic installation involves the observer-participant in an immersive mixed reality environment in which he sees a life-size, realistic, virtual human character sitting in a real chair; he/she can move freely within a relatively large designated area while still being tracked by the system. This installation is an amalgamation of art and technology – the content development, model design, animation, augmented/mixed reality, computer vision and inertial sensing.

Future works include:

- Developing interaction techniques which allows the observer-participants to interact with the virtual subjects/objects in a more intuitive way;
- Tackling of the occlusion problem whereby the real and virtual objects can occlude each other (this is possible if the geometrical models and exact coordinates of the real objects are known);
- Incorporating artificial intelligence in the virtual human;
- Researching into natural feature tracking techniques which could potentially eliminate the use of IR beacons and inertial sensor
- Photorealistic and stereo rendering can also be explored in order to provide users truly immersive and realistic entertainment.

Other than for cultural and historical installation, this work has the potential of being further developed for the purpose of education, art, entertainment and tourism promotion.

References

ARTOOLKIT http://www.hitl.washington.edu/artoolkit

CAARLS, J., JONKER, P.P., PERSA, S. Sensor Fusionfor Augmented Reality. *EUSAI 2003:* 160-176

HOL, J.D. SCHON, T.B., GUSTAFSSON, F., SLYCKE, P.J. 2006. Sensor fusion for augmented reality. In *Proceedings of 9th International Conference on Information Fusion.*

KATO, H., BILLINGHURST, M. 1999. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*. October, San Francisco, USA

KSATRIA GAMEWORKS. http://www.ksatria.com

MXRTOOLKIT http://sourceforge.net/projects/mxrtoolkit

NGUYEN, T.H.D., QUI, T.C.T., XU, K., CHEOK, A.D., TEO, S.L., ZHOU, Z.Y., ALLAWAARACHCHI, A., LEE, S.P., LIU, W., TEO, H.S., THANG, L.N., LI, Y., and KATO, H. 2005. Real Time 3D Human Capture System for Mixed-Reality Art and Entertainment, *IEEE Transaction on Visualization and Computer Graphics (TVCG)*, 11, 6 (November-December 2005), 706 - 721.

SLABAUGH, G., SCHAFER, R., LEVINGSTON, M. 2001. Optimal ray intersection for computing 3D points from N-view correspondences.

SONG, M., ELIAS, T., MARTINOVIC, I., MUELLER-WITTIG, W., CHAN, T.K.Y., 2004. Digital Heritage Application as an Edutainment Tool. In *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry*

SPARACINO, F., WREN, C., DAVENPORT, G., PENTLAND, A. Augmented performance in dance and theater. *In International Dance and Technology 99*, ASU, Tempe, Arizona, February 1999.

UEMATSU, Y., SAITO, H. 2007. Improvement of accuracy for 2D marker-based tracking using particle filter. In *Proceedings of 17th International Conference on Artificial Reality and Telexistence.*

WII REMOTE http://wiibrew.org/wiki/Wii_Remote

WII YOURSELF. http://wiiyourself.gl.tter.org/