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

William Russell Pensyl, Tran C.T. Qui, Pei Fang Hsin, Shang Ping Lee and Daniel K. Jernigan

Nanyang Technological University, Singapore

***Abstract*— We have developed a system which enables us to track participant-observers accurately in a large area for the purpose of immersing them in a mixed reality environment. This system is robust even under uncompromising lighting conditions. Accurate tracking of the observer's spatial and orientation point of view is achieved by using hybrid inertial sensors and computer vision techniques. We demonstrate our results by presenting a 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 even while still being able to see the avatar from various angles. The project installation provides an exciting way for cultural and historical narratives to be presented vividly in the real present world.**

***Index Terms*—cultural and historical installation, hybrid vision and inertial sensor, localization, mixed reality**

I. INTRODUCTION

Tracking black-and-white fiducial markers [1] has always been the conventional way in augmented reality (AR) or mixed reality for 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 real world coordinates and/or the marker's coordinates. However, this technique suffers under uncompromising lighting condition, which can cause jittering 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 [2]. In addition to this, for large area tracking, big markers or several markers would have to be dispersed in the environment, which can be disruptive of the visual aesthetic.

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 sensors. 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 use 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 systems. Furthermore, an inertial sensor consisting of 6DOF is attached to the HMD to detect the rotation and movement of user viewpoints.

Previous work [3] tracked humans in an environment "whose only requirements are good, constant lighting and an unmoving background". In [4], 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.

The work in [5] provided excellent accounts and experimental results of hybrid inertial and vision tracking for augmented reality registration. The sensitivities of orientation tracking error were quantitatively analyzed, natural feature tracking and motion-based registration method that automatically computed the orientation transformation (among different coordinate systems) was presented. Our work is different in that the tracking cameras (Wii Remotes) are actually only tracking IR light point source; and they were statically mounted on the wall instead of attached to the observer's head.

II. BACKGROUND

The motivation for developing this project is to allow people to experience pseudo-historical events impressed over present day real world environments (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 the famous people who frequented the bar in the early 20th 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 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 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 that can accommodate compromising lighting conditions and large open spaces.

One of the early works investigating AR technology for cultural heritage was [6]. It gave a good overview of the AR and virtual reality (VR) technology and applications in several areas, in particular the cultural heritage field. It pointed out that the problem with working with large spaces such as museums is the technology of tracking systems. We have addressed the indoor tracking issue in this work.

Reference [7] developed a VR system for digital heritage application. Their system allowed users to navigate through a virtual heritage site, with the aid of a virtual tour guide. However, it did not track the user's whereabouts; the navigation was achieved by users pointing a device (a brush) on an interactive screen. Our system differs in a few ways: it is an augmented and mixed reality system which allows users to physically walk around within the area of the installation while yet being able to see the virtual (but vividly realistic) human sitting in a real world environment.

Reference [8] proposed a mobile, hand-held display guide for use in the context of visiting museums. It also attempted to address the concerns of latency, and proper alignment and registration of digital objects to the real scene, and acknowledged that these were crucial to the acceptance and success of the system and that unfortunately no existing approach completely satisfied these requirements.

Our work prefigures new entertainment forms we refer to as interactive entertainment. In this new form narrative, animation, and film based presentations 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 while placing the viewer behind the "fourth wall" and immersing them in ever new and experiential ways into the performance installation.

III. SYSTEM DESCRIPTION

3.1 Review Stage

Fig. 1 shows the system setup. The observer wears a HMD and can move freely around within the designated area. There is an empty chair, a table and other physical artifacts in the demo area. As the observer looks 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.

3.2 Low-cost Vision-based Tracking of Head Position

Conventional ARToolkit [9] or MXRToolkit [10] markers are not suitable for vision-based tracking in large area, uncompromising lighting environments. Jittering and loss of tracking due to the lighting conditions seriously hampers accurate tracking, adversely impacting the audience's aesthetic

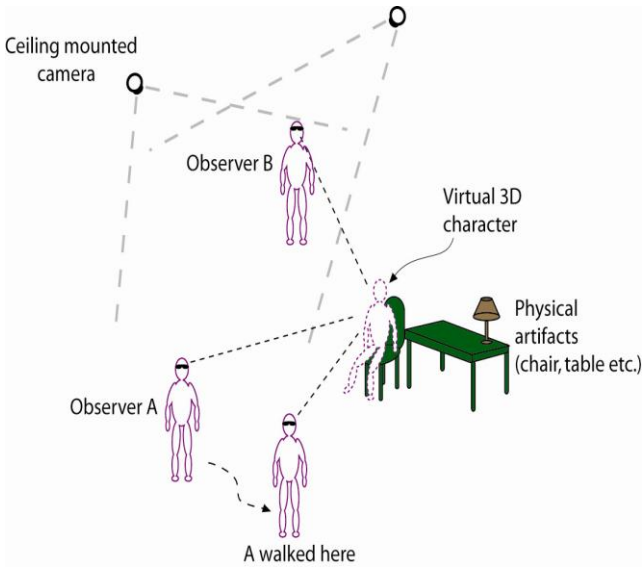

Fig. 1. Setup of the large area robust hybrid tracking system. The chair, table and lamp are physical objects in the real world. The "human" in dashed-line and sitting in the chair is a 3D avatar.

experience. In view of this, we have devised 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 Remotes 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 [11]. 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 images – the coordinates of the tracked points – are sent to the host) and wireless. The host computer is relieved from processing the raw image; and an optimal triangulation technique [12] is all that is needed to obtain the depth information of the IR LED.

3.3 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 apply sophisticated sensor fusion algorithms (with Kalman filter) [13][14] which combine with the vision-based positioning reading to obtain more accurate results.

The filter we are using employs a velocity model with 6 parameters, 3 parameters for camera position and 3 parameters for velocity.

$$x_k = \begin{bmatrix} X & Y & Z & V_x & V_y & V_z \end{bmatrix}^T$$

At each time step *(k)*, the update from the previous step *(k-1)* is computed as:

$$x_k = Fx_{k-1} + Ga_{k-1}$$

where:

$$F = \begin{bmatrix} I_3 & \Delta t.I_3 \\ 0 & I_3 \end{bmatrix},\ G = \begin{bmatrix} \frac{\Delta t^2}{2}.I_3 \\ \Delta t.I_3 \end{bmatrix}$$

a_{k-1}: values from the accelerometer sensor

The covariance matrix P of the filter state is updated according to the dynamic model as:

$$P_{k|k-1} = F.P_{k-1|k-1}.F^T + Q_{k-1}$$
$$Q_{k-1} = \sigma_a{}^2 GG^T$$

(σ_a : standard deviation of acceleration distribution)

This x_k, computed using information from the inertial sensor, will be compared with the z_k, the camera position computed from the triangulation algorithm.

$$y_k = z_k - Hx_k$$

where:

$$H = \begin{bmatrix} I_3 & 0_3 \end{bmatrix}$$

With the difference y_k, the final $\hat{x}_k$ will be computed:

$$\hat{x}_k = x_k + K_k.y_k$$

Where:

$$K_k = P_{k|k-1}.H^T.S^{-1}{}_k \quad \textit{(Kalman gain)}$$

$$S_k = H.P_{k|k-1}.H^T.\sigma_z{}^2.I^3$$

(S_k : Innovation covariance)

(σ_z : standard deviation of noise distribution)

The calculation of $\hat{x}_k$ as above has fused information from both inertial sensors and vision based techniques, thus provides better tracking results. Currently, we only used this filter for tracking camera position, as we can get only head location fro one IR LED. Further research will be looking into fusing head rotation as well.

3.4 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 a 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.

Fig. 2 shows the animation sequence of the virtual human. As can been seen, the model has high quality textures and proportionate skeleton. The animation was created by obtaining the body postures through 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, meaning that essentially no computer-generated in-betweens are used. This further enhances the natural motion of the animation.

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

Fig. 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.

3.5 3D Model, Rendering and Animation of the Virtual Human

Fig. 3 shows the block diagram of the system. The peripheral hardware is shaded in blue.

We use the kjAPI game engine [15], developed by Ksatria Gameworks, to render visual content and animation. 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 is used. In spite of using only one workstation, the system can operate at 30 frames per second and provides users a fluidly interactive and realistic entertainment.

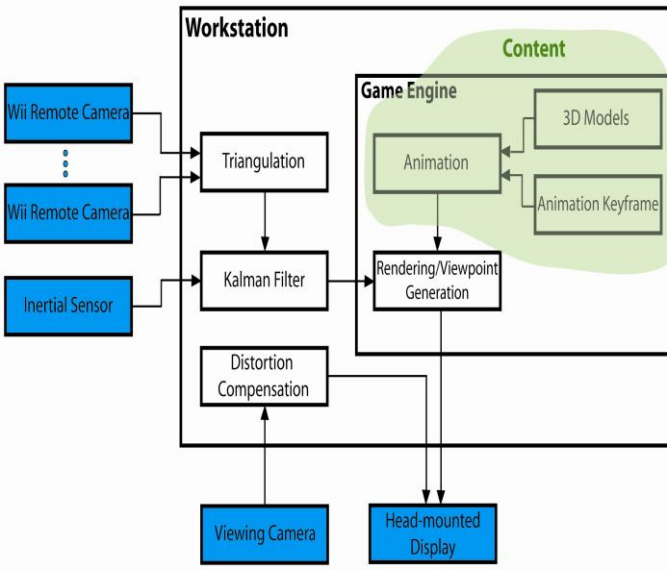

Fig. 3. Block diagram of the system.

The head-mounted display comes with binocular vision function. However, since there is only one viewing camera attached to the HMD, the binocular vision function is not used.

IV. INSTALLATION

The project was installed in the DemoGraphics in-conjunction exhibition of the International Symposium on Electronic Art at the Inner Gallery of Nanyang Technological University. The designated area was about 4m x 4m. This space could have been larger or smaller depending on the camera lens and the strength of the IR LED. As shown in Fig. 4 and Fig. 5, the system continued to function even with a mix of studio lighting and spill over from various outside lighting sources.

Fig. 5 depicts the potential for using mixed and augmented reality technologies in cultural and historical installations – it breaks the time-and-space barrier between present day people and all things of the past – the real present day participant standing beside the virtual human of the past.

Other participants, not equipped with the viewing camera and tracking devices, would be able to see exactly what the observer with HMD is seeing – through a desktop LCD display (Fig. 4).

Fig. 4. The installation of the Long Bar. The observer wore the HMD and looked at the table and chair. With naked eyes, observer-participants only saw the physical objects. The view seen in the observers HMD was duplicated in a monitor for other viewer to peripherally experience the scene.

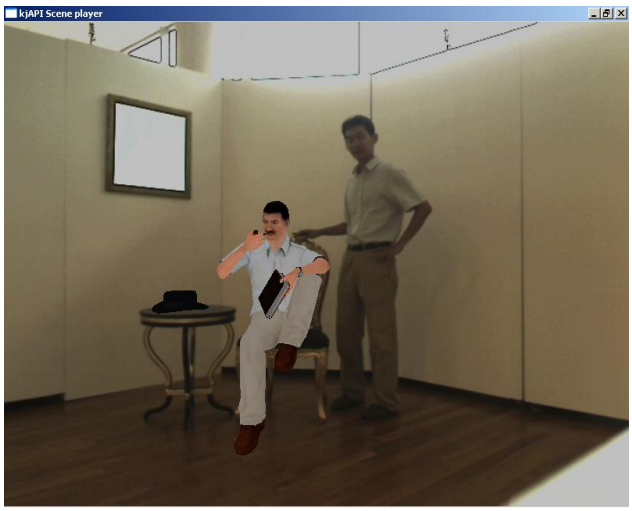

Fig. 5. This is the view through the observer's HMD. It is as if the live participant is resting his hand on the shoulder of the virtual human. The lighting is dimmed as compared to that of Fig. 4 but the tracking still works robustly.

V. CONCLUSION AND FUTURE WORKS

We have developed a highly robust, low-cost, hybrid tracking system for use in augmented reality applications. 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

[1] H. Kato, M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system, in *Proceedings of the 2nd International Workshops on Augmented Reality (IWAR 99)*, San Francisco, 1999.

[2] Y. Uematsu, H. Saito. Improvement of accuracy for 2D marker-based tracking using particle filter, in *Proceedings of 17th International Conference on Artificial Reality and Telexistence*, 2007.

[3] F. Sparacino, C. Wren, G. Davenport, A. Pentland. Augmented performance in dance and theater. In *International Dance and Technology*, ASU, Tempe, Arizona, 1999.

[4] T.H.D. Nguyen, T.C.T. Qui, K. Xu, A.D. Cheok, S.L. Teo, Z.Y. Zhou et. al. 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.

[5] S. You, U. Neumann, R. Azuma. Hybrid Inertial and Vision Tracking for Augmented Reality Registration. In *Proceedings of the IEEE Virtual Reality* (March 13-17, 1999). VR. IEEE Computer Society, Washington, DC, 260

[6] B.A. Brogni, C.A. Avizzano, C. Evangelista, M. Bergamasco, P. Percro. Technological approach for cultural heritage: augmented reality. In *Proceedings of the 8th IEEE International Workshop on Robot and human interaction.* pp. 206-212, 1999

[7] M. Song, T. Elias, I. Martinovic, W. Mueller-Wittig, T.K.Y. Chan. 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*, 2004

[8] A. Damala, I. Marchal, P. Houlier. 2007. Merging augmented reality based features in mobile multimedia museum guides. In *Anticipating the Future of the Cultural Past, CIPA Conference 2007*, 1-6 October 2007, Athens, Greece, pp. 259-264.

[9] ARTookit. http://www.hitl.washington.edu/artoolkit

[10] MXRToolkit. http://sourceforge.net/projects/mxrtoolkit

[11] Wii Remote. http://wiibrew.org/wiki/Wii_Remote

[12] G. Slabaugh, R. Schafer, M. Levingston. Optimal ray intersection for computing 3D points from N-view correspondences. 2001

[13] J. Caarls, P.P. Jonker, S. Persa. Sensor Fusion for Augmented Reality. *EUSAI 2003:* 160-176

[14] J.D. Hol, T.B. Schon, F. Gustafsson, P.J. Slycke. Sensor fusion for augmented reality. In *Proceedings of 9th International Conference on Information Fusion.*2006

[15] Ksatria Gameworks. http://www.ksatria.com

William Russell Pensyl is a American media artist and designer working in Singapore. He maintains a strategic focus on communication, narrative, and user centric design processes for interactive media and communication media. His work is diverse, including projects for film, broadcast, interactive multimedia, internet development, art and installation. Pensyl's professional practice in digital media includes designing and creating films and interactive media based projects for Fortune 500 companies such as Apple Computer, IBM, Motorola, Kodak, Adobe Systems, Sony, Disney, Pearson Education, JCPenney, WebTV, 3Com, PalmPilot, American Airlines, Lucent Technologies and others.

Pensyl's current work concerns the creation of location based entertainment using virtual characters set up real-world environments using new tracking technologies in mixed and augmented reality. Pensyl's exhibition credits include: Installation of a mixed reality experiences in SIGGRPAH ASIA, Art Gallery. the international DAT exhibition in Singapore, an experimental theatre production, "Everyman, The Ultimate Commodity," staged in the Fringe Toronto Theatre Festival, a interactive new media installation in the Second International Science and Art Exhibition in Beijing, China; the Shang Hai Biennial in 2004; an innovative media installation, Journey to the Oceans of the World in the Art Gallery at SIGGRAPH 2002 in the United States and Art on the Net, 2002 - 9.11, an international exhibit hosted by Machida City Museum of Graphics Arts, Tokyo, Japan. Pensyl has won several distinguished awards for art direction in media and his film and media presentations have been included at Comdex, MacWorld, National Association of Broadcasters, SIGGRAPH and TeleCom in Europe.

Pensyl is currently Associate Chair Academic in the School of Art, Design & Media, Director of the Interaction and Entertainment Research Center and Co-Director of the Institute for Media Innovation, Pensyl was the founding Director of Computer Animation and Chair of the Department of Digital Art and Design at Peking University, where he designed and implemented the first comprehensive Graduate Degree program in Digital Art and Media at a top university in the People's Republic of China. Pensyl's previous posts were an Associate Professorship in Digital Animation and Interactive Media at the William Paterson University of New Jersey and an Associate Professorship in Communication Design and Computer Art at the University of North Texas.

Tran Cong Thien Qui was born and raised in Vietnam. He received the B.Eng. degree (University Gold Medal) in Information Technology from the Ho Chi Minh City University of Technology, Vietnam in 2003 and the M. Eng. degree in Electrical and Computer Engineering from National University of Singapore in 2006.

He is currently a Research Associate at the Interaction and Entertainment Research Center (IERC), Nanyang Technological University (NTU), Singapore. His research output has been published in numerous academic journals, international conferences and exhibitions. His research interests include computer vision, computer graphics, human – computer interaction and game development.

Pei Fang Hsin currently holds a position as a conceptual designer in the Interaction & Entertainment Research Centre at Nanyang Technological University. She has done both in-house and freelance design projects for 9 years. Her focus is on concept development, visual communication and presentation utilising 2D and 3D designed elements.

Her professional experience is international: Taipei, Taiwan, Shanghai, China., San Francisco California, New York, NY, Dallas, Texas, USA and more recently in Singapore. The work covers a range of design for print media, digital media and interior design giving her opportunities to design projects across multiple cultures and disciplines.

Shang Ping Lee was born in Kuching, Malaysia. He obtained his B.Eng (Mechanical & Production Engineering) and M.Sc (Mechatronics) at the Nanyang Technological University in 1997, and the National University of Singapore in 2001 respectively.

He worked at IBM Singapore Pte. Ltd from 1997 – 1999, and then at SGP Global Technologies in 2000 before starting his research career at the National University of Singapore as a research engineer in 2001. He then joined the Interaction and Entertainment Research Centre at Nanyang Technological University in 2005. His technical skill sets include DSP, embedded system, printed circuit board design, FPGA/VHDL and firmware programming. His research interests are wearable technologies, sensors, and human-computer interface (HCI).

Daniel K. Jernigan is an assistant professor of English at Nanyang Technological University, Singapore. He studied at Purdue University, Indiana, receiving the PhD in 2002. His interests include drama and theatre studies, postmodernism, playwriting, and science studies. Dr. Jernigan's essays on Caryl Churchill and Tom Stoppard have been published in *Modern Drama*, *Comparative Drama*, and *Text and Presentation*. He is editor of *Drama and the Postmodern: Assessing the Limits of Metatheatre* (Cambria Press, 2008). He is also an aspiring playwright, whose plays have been produced in Singapore and Toronto and published in *The Massachusetts Review*.