Eurographics Italian Chapter Conference (2011)
A. F Abate, M. Nappi, and G. Tortora (Editors)

Crime Scene Interpretation Through an Augmented Reality Environment

A. Casanova[1], M. De Marsico[2], S. Ricciardi[3]

[1]University of Cagliari - [2]University of Roma, La Sapienza - [3]VRLab, University of Salerno, Italy

Abstract

Despite its potential advantages, gesture based interface usage is currently rather limited due to operational and practical issues, while most proposals aim at replacing mouse and keyboard functionalities for medical/surgical applications. This paper presents a crime scene interpretation framework which combines augmented reality visual paradigm and gesture based interaction to provide a new generation of detectives with interactive visualization and manipulation of virtual exhibits while seeing the real environment. The idea is to augment the exploration of the crime scene by means of a see-through head mounted display, exploiting a small set of simple (user-wise) gestures and the visual interface to enable a wider set of commands and functionalities, improving both the efficacy and the accuracy of user-system interaction. The proposed system allow the user to freely position virtual replicas of real object to interactively build visual hypothesis about the crime under investigation, or even to set virtual landmarks which can be used to take distance/angular measurements. All these action can be performed without mouse and keyboard but simply through intuitive gestures.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities. H.5.2 [User Interfaces] Input devices and strategies. I.3.6 [Methodology and techniques] Interaction techniques.

1. Introduction

Over the last decade, the use of VR (Virtual Reality) in forensic has become more and more common, as a practical and impactful way to visualize and communicate hypotheses and fact reconstructions according to exhibits collected on crime scene or even to witness testimonies.

The growing use of this technology is partly due to the increased visual processing power of last generation computers and partly to the reduction in cost and time needed to build a VR based forensic simulation.

On the other side, AR (Augmented Reality) technologies have improved as well and both motion tracking systems and see-through HMDs (Haed Mounted Displays) are today available in a much more affordable range of costs, thus enabling the visualization of augmented contents in many unexplored applicative fields [ABF*01]. Among these, crime scene investigation represents an interesting domain in which to transfer many techniques and results reported in AR scientific literature, as there are many reasons to explore not just a virtually reconstructed crime scene [ADL*08], but to augment, instead, the real crime scene with virtual contents useful to the development of the investigation process.

To this regard, a number of well established theories assert the possible contributions which may be provided to the analysis of investigative hypotheses, by allowing to visually verify position and eventual displacement of exhibits, bodies, weapons or crime dynamics.

The aim of this study is to propose a novel crime scene interpretation paradigm by using augmented reality and gesture based interaction techniques to foster an innovative investigation methodology. The paper describes the hardware and software technologies involved in setting up such a mixed reality environment.

We believe that AR based exploration of a crime scene, in which gesture based interaction allows the investigator to visually interact with the surrounding augmented environment by positioning or repositioning objects relevant for hypothesis formulation or testing, may represent a powerful tool in investigators' arsenal.

The rest of this paper is organized as follows. Related works are exposed in section 2, while the proposed system is described in section 3, and briefly discussed in section 4. The paper concludes in section 5.

DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2011/029-033

2. Related works

AR technology is becoming familiar in many different applicative scenarios.

In the automotive context, a framework for instance visualisation and modification of car body curvature and engine layout has been proposed [FAS02]. On the same line, a simulation where one can open the door of a virtual concept car and experience the interior, dash board layout and interface design for usability testing has been presented [TSK*05]. In the assembly context, Pentenrieder et al. [PBDM07] shows how leading automotive company use AR during car assembly to visualize interfering edges, plan production lines, compare variance and verify parts [BKFT00]. An extra benefit of augmented assembly and construction is the possibility to monitor and schedule individual progress in order to manage large complex construction projects [DFG*05]. An example by Feiner et al. [FMH99] generates overview renderings of the entire construction scene while workers use their HMD to see which strut is to be placed where in a space-frame structure. Distributed interaction on construction is further studied by Olwal and Feiner [OF04]. AR is also proving useful in the area of maintenance, for industrial plant inspection [KBCW03], for electrical troubleshooting [FRI02] and for training/learning [BRI91], though (to the authors' best knowledge) AR does not result utilized in forensics yet.

On the other hand gesture based interfaces are surely not new in human-computer-interaction literature. They were originally presented as the most natural way to navigate through and interact with virtual reality environments, generated by military simulators or by entertainment applications.

More recently, thanks to technological progress and cost reduction, they have been proposed for advanced interaction in a broader spectrum of applicative fields, ranging from industrial processes to medical applications [KHLE97]. In this latter field the researches on gesture based interaction are often focused on the replacement of conventional input devices like keyboard and mouse with a contact-less interface for sterile operating room interventions. For this kind of applications a visual approach to gesture recognition [KS98] is preferred because it does not require the user to wear any specific device as the recognition is based on video acquisition and processing of gestures. Along this line of research other authors [GFGB04] exploit stereo cameras to capture both color and depth info to achieve reliable, high-speed hand detection and tracking within a user-specified workspace, and to interpret hand gestures as mouse commands (pointer movement and button presses).

This paper proposes a mixed reality and gesture based environment, targeted to interactive exploration and analysis of a crime scene. The proposed framework combines one hand and two hands gestures with an augmented context adaptive graphic interface [PTB*02], to interactively display virtual objects augmenting the crime scene while navigating it.

3. Overall system's architecture

The proposed framework is schematically depicted in Fig. 1. and can be operatively summarized as follows. After a brief calibration session, which is needed to initialize gloves sensors and check the trackers, the system enters in operating mode. Two separate data channels, respectively for the left and right hand, are pre-processed and then both feed the Gesture Recognition Engine. This module analyzes posture tokens outputted by datagloves, combined with motion data captured by the tracking system As a result of this process a (one-hand or two-hands) gesture is recognized and passed to the Interaction Engine, which eventually selects a function of the virtual keyboard or translates the gesture in a transformation (rotation/traslation) of the 3D model sent to the AR Visual Engine. Here the head tracking data are exploited to co-register the virtual contents to the real scene as seen through the optical see-through HMD. Therefore, two renderings (left and right) of the augmenting objects are calculated and coherently displayed by the HMD. The main hardware and software components of system architecture are described in the following sub sections 3.1 to 3.3.

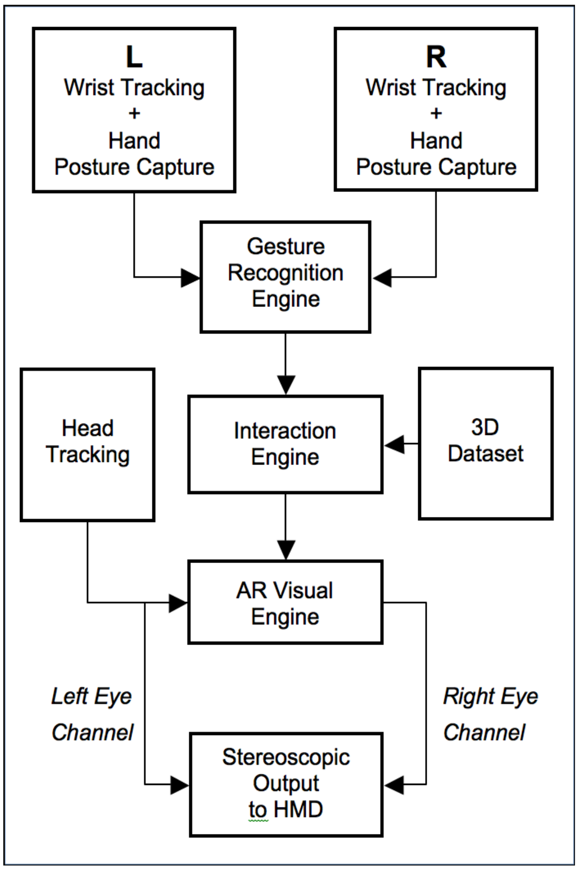

Figure 1. *Overall view of the proposed framework.*

3.1 Motion tracking and gesture recognition

For this particular application, head tracking should be portable and scalable, as it is supposed to be deployed on crime scene location, even if we restricted the system's applicative field to indoor use only. Head tracking is based on magneticl motion tracking as it provides good accuracy in a operative field which can grow adding more magnetic field generators.

The left/right input module is responsible for user's hands tracking within 3D space and hand posture acquisition. Here an accurate and reliable capture technique can be exploited, based on wireless instrumented gloves and magnetic tracking devices. Such choice simplifies the gesture recognition stage (e.g inter-hands and inter-fingers occlusions are not an issue anymore) as each single finger has individual sensors for flexion and abduction which are unaffected by any other fingers.

More in detail, left and right hand posture acquisition is performed via a couple of wireless 5DT Dataglove 14 ultra, featuring fourteen channels for finger flexion and abduction measurement, with 12 bit of sampling resolution each. Additionally, a binary (open/closed) value for each finger, resulting from the comparison of the normalized joint flexion value to a threshold, leads to 24 (four fingers are considered excluding the thumb) different combinations or postures. Though it is possible to access a much more accurate finger status representation via each sensor's raw values, for the posture recognition this built in classification is more appropriate as it simplifies both user training (partially flexed fingers do not compromise posture recognition) and the gesture recognition module design.

The four hand postures used in this study are: *fist*; *index finger point*; *not index finger point*; *flat hand*.

They have been selected as they are the simplest to perform for most users and among the most used in natural interaction. As datagloves do not provide any spatial information, the system relies on Polhemus Liberty Latus magnetic motion tracking hardware, with six degrees-of-freedom, to detect head and wrists position in 3D space and their rotation on three axis (yaw, pitch and roll). A preprocessing applied to each of six channels (for each hand) filters capture noise by means of a high frequency cut and a temporal average of sampled values. Then both left and right data streams are outputted, each including basic postures captured and positional/rotational information.

At this stage a specific software module checks for particular posture tokens which trigger associated interaction activities. The analysis is performed by means of a Gesture Recognition Engine based on timed automata [AD94], able to detect one-hand, two-hands and timed posture patterns which are associated to manipulation functions. Timed automata are labeled transition systems used to model the behavior of components in real-time systems over time. State-transition graphs are annotated with timing constraints. A timed automaton performs time-passage actions, in addition to ordinary input, output and internal actions. It accepts timed words, i.e. infinite sequences in which a real-valued time of occurrence is associated with each symbol. Timed automata so provide a feature that classical finite automata do not address in any way, namely timing. Embedding time allows changing the status of involved entities according to time-based events to enhance the quality of user-system interaction. The aim here is to augment the basic one-hand postures through timed patterns or via a combination of left and right hand for a simple yet more powerful interaction. The use of timed automata offers a further key benefit for the proposed architecture, as it enables the application designer to formally verify the interaction model by means of well established model checking procedures. In the proposed framework, eight gestures are used as considered, two of which are defined through a two hand combination of basic postures, while other two are defined by a timed sequence of basic postures (for instance fist-flat_hand-fist, or double pointing). Recognized gestures are represented by a vector including gesture index, first hand x-y-z spatial coordinates, first hand yaw-pitch-roll angles, second hand x-y-z spatial coordinates, and second hand yaw-pitch-roll angles.

3.2 Gesture based interaction

Once a valid gesture has been recognized the corresponding gesture vector is passed to the interaction engine, which exploits the same timed automata based design of recognition engine, and is responsible for any visual interaction allowed by the system, translating user's gestures into actions. Gestures are evaluated according to the current interaction status, so that the same gesture may trigger different actions in different operative contexts (rotation, measurements, landmark assignment, etc).

Operational modes and manipulation functions are selected via a virtual interface, which is displayed within the field of view as a frame surrounding the 3D content, and including textual info related to the ongoing operations. The layout of such interface is visually perceived as a floating object, thanks to stereoscopic rendering. It is displayed in a close at hand position along the depth of visual field, according to a calibration procedure in which the user touches with the finger a sequence of small targets at various depths, thus allowing an adaptation of the parameters regulating the stereo effect.

One of the main aims of interface design is to minimize the number of gestures required to operate it, yet preserving familiar interaction patterns. For this reason a gesture adaptation of the classical point and click interaction paradigm is adopted: selection is triggered by hitting an active area by index fingertip, an action or a confirm is triggered by double hitting, a cancel/escape command is triggered by a fist-flat_hand-fist sequence.

Though only one-hand gestures are used to operate the interface, an experienced user could exploit both hands to operate in a faster and more comfortable way (for instance, typing characters in a text field by both hands results in a much faster operation than via a mouse based selection and would not require a physical keyboard). Visual and acoustical feedbacks are provided to confirm the “pressure” of a key or the acknowledgment of a particular command, thus reducing wrong operations. If required, interface layout can be hidden at any time via a gesture toggle. At the moment, only a small set of functions has been implemented, allowing to move/rotate/ an object, to place virtual landmarks over the real scene and to take distance measurements between landmarks.

For any task involving positioning in 3D space, a precise calculation of actual finger positions is performed through a combination of a forward-kinematics applied to a 3D parametric hand model, which is adapted to the real user’s hand measures during a calibration session. This setup is due only one time and may be saved and retrieved when the system is initialized. In this case the raw flexion values are exploited for each finger.

3.3 Scene augmentation

3D models and the virtual interface are both processed by the visualization engine, also responsible for AR related real time transformations and for the stereo rendering of 3D content. The engine is built on the Quest3D graphics toolkit and DirectX API. To provide the AR environment, the Visualization Engine exploits user’s biometrics such as head’s position and orientation, then processes these data to transform the virtual content as seen from user’s point of view and coherently to a 3d model of surrounding environment, a crucial task referred as 3D registration.

Actually, any AR architecture requires a precise registration of real and virtual objects. In other words, the objects in the real and virtual world must be properly aligned with respect to each other, or the illusion that the two worlds coexist will be compromised.

To achieve this not trivial goal two main requirements have to be satisfied: a) the position and orientation of user’s head have to be precisely tracked at a high sample rate; b) the physical world or at least the objects relevant to the application has to be precisely measured in the same 3D space in which the user operates.

At runtime, two rendering cameras (one for each eye) are built, matching the exact position/orientation of user’s eyes, transforming each vertex of each virtual object to be displayed onto the real scene accordingly. Two renderings (left and right) are then calculated and coherently displayed through an optical see-through Head Mounted Display (a Cybermind Visette SXGA), which is an helmet working by placing optical combiners in front of the user's eyes (see Fig. 3). These combiners are partially transmissive, so that the user can look directly through them to see the real world. The combiners are also partially reflective, so that the user sees virtual images bounced off the combiners from head-mounted LCD monitors.

The rendering engine has been tailored to optical see-through HMD, but it can be easily adapted to video see-through displays. Eventually, a selective culling of a virtual object may be performed whereas it is partially or totally behind a real object, but for the applications showcased in the presented case study this technique (and the overhead required to accurately model the real environment) could not be necessary.

4. Discussion

The framework described so far is only a part of a wider project which aims at enhancing and extending crime scene investigation and interpretation skill. As it is still in a early stage of development, the first concern is to assess if the proposed approach to crime scene augmentation and virtual object interaction is sustainable (technology wise) and compatible to the most recent investigation methodology.

Today, modern investigation techniques are often based on crime scene 3D capture and analysis by means of conventional mouse and keyboard-based software tools. As the three dimensional data generated by these approaches can be quite complex, problems may easily arise, related to the way these data are accessible, and to how complex is interacting with them through a real-time virtual environment. Indeed, though these digital replicas can be accurately reconstructed from reality, the investigator’s presence on the real crime scene can not be matched by a virtual exploration of it.

The challenge is therefore to combine the capabilities offered by synthetic images with a more natural way of observing the crime’s location and evidences. For this reason the “natural interaction” requirement is very important and the choice of a more complex and costly gesture based interface makes sense. Indeed, operating on a 3D volume or volume requires a more powerful way to specify actions or locations in space which is not well addressed by usual 2D-based interaction paradigm, involving conventional interfaces. At the same time, many conventional functions have to be accessible and easy to use in a way similar to that possible with common devices (mouse and keyboard).

Figure 1 – *User wearing left and right hand 5DT- 14 dataglove ultra, plus head, left hand and right hand 6DOF motion trackers. A Cybermind see-through HMD is also worn to visualize the virtual contents onto the surrounding scene.*

5. Conclusions

In this paper a framework for crime scene interpretation through an augmented reality environment has been presented. The potential advantages of this approach may be summarized as follows: it provides a powerful tool for visual hypothesis formulation and verification, by means of either a physical presence on the crime scene or an interactive augmentation of it through a natural interface.

The proposed gesture-based interface approach exploits one-hand and two-hands timed interaction patterns, combined with a virtual floating interface. The preliminary tests conducted seem promising, as they show an user-wise perceptible improvement in performing 3D interaction tasks like the free positioning of virtual objects within the real environment or the selection of a particular spot or the distance measurement between features.

As the work is in a early stage, more functions have to be implemented and complete and accurate experiments are required to measure the advantages provided to the real investigative practice.

References

[ABF*01] Azuma, R. Baillot, Y. Behringer, R. Feiner, S. Julier, S. MacIntyre, B. (2001) Recent advances in augmented reality, IEEE Computer Graphics and Applications, Volume: 21, Issue: 6, 2001, pp.34-47

[AD94] Alur, R and Dill, D. L. (1994). A theory of timed automata. Journal of Theoretical Computer Science, 126(2):183–235.

[ADL*08] A. F. Abate, M. De Marsico, S. Levialdi, V. Mastronardi, S. Ricciardi, G. Tortora. Gesture Based Interface for Crime Scene Analysis: A Proposal, Proceedings of VRSAL 2008 Conference, Perugia, Springer, pp 143-154, 2008

[BKFT00] S. Balcisoy, M. Kallmann, P. Fua, D. Thalmann (2000). A framework for rapid evaluation of prototypes with augmented reality, Proceedings of the ACM symposium on Virtual reality software and technology, Seoul, Korea, pp. 61 – 66, 2000

[BRI91] M. Bricken (1991). Virtual reality learning environments: potentials and challenger, ACM Computer Graphics, Volume 25 , Issue 3, pp. 178 – 184, 1991

[DFG*05] W. Dangelmaier, M. Fischer, J. Gausemeier, M. Grafe, C. Matysczok and B. Mueck (2005). Virtual and augmented reality support for discrete manufacturing system simulation. Computers in Industry, Volume 56, Issue 4, May 2005, Pages 371-383

[FAS02] M. Fiorentino, R. de Amicis, G. Monno, and A. Stork. Spacedesign: A mixed reality workspace for aesthetic industrial design. International Symposium on Mixed and Augmented Reality (ISMAR'02), 2002, pp. 86–318.

[FMH99] S. Feiner, B. MacIntyre, and T. Höllerer. Wearing it out: First steps toward mobile augmented reality systems. Ohmsha (Tokyo)-Springer Verlag, Berlin, 1999, pp. 363–377.

[FRI02] W. Friedrich, “ARVIKA-Augmented Reality for Development, Production and Service,” Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR '02), 2002, pp. 3–4.

[GFGB04] C. Graetzel, T. Fong, S. Grange, and C. Baur. (2004). A Non-Contact Mouse for Surgeon-Computer Interaction. Technology and Health Care, IOS Press, vol. 12, no. 3, pp. 245-257.

[KBCW03] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. ISMAR 2003, pp. 133–141

[KHLE97] C. Krapichler, M. Haubner, A. Lösch, and K. Englmeier, (1997) “Human-Machine Interface for Medical Image Analysis and Visualization in Virtual Environments”, IEEE conference on Acoustics, Speech and Signal Processing, ICASSP-97. Vol 4, pp. 21-24.

[KS98] Kohler, M. and Schroter, S. (1998). A Survey of Video-based Gesture Recognition - Stereo and Mono Systems. Technical Report 693, Informatik VII, University of Dortmund.

[OF04] A. Olwal and S. Feiner. Unit: modular development of distributed interaction techniques for highly interactive user interfaces. ACM New York, NY, USA ©2004, pp. 131–138.

[PBDM07] K. Pentenrieder, C. Bade, F. Doil, and P. Meier. Augmented reality-based factory planning -an application tailored to industrial needs. In ISMAR’07: Proc. 6th Int’l Symp. on Mixed and Augmented Reality, pp. 1–9, Nara, Japan, Nov. 13-16 2007. IEEE CS Press. ISBN 978-1-4244-1749-0.

[PTB*02] Poupyrev, I., Tan, D.S., Billinghurst, M., Kato, H., Regenbrecht, H., Tetsutani, N. (2002) Developing a generic augmented-reality interface, Computer, Volume: 35, Issue: 3, 2002, pp. 44-50

[TSK*05] M. Tönnis, C. Sandor, G. Klinker, C. Lange, and H. Bubb. Experimental evaluation of an augmented reality visualization for directing a car driver‟s attention. ISMAR '05 Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 56–59.