A Pervasive Gesture-Driven Augmented Reality Prototype using Wireless Sensor Body Area Networks

Peter Barrie
Lecturer
Glasgow Caledonian University
70 Cowcaddens Rd.
Glasgow G4 0BA
UK
+44 141 331 3025
Peter.Barrie@gcal.ac.uk

Andreas Komninos
Lecturer
Glasgow Caledonian University
70 Cowcaddens Rd.
Glasgow G4 0BA
UK
+44 141 331 3095
Andreas.Komninos@gcal.ac.uk

Oleksii Mandrychenko
MSc student
Glasgow Caledonian University
70 Cowcaddens Road
Glasgow G4 0BA
UK
+3809798400068
oleksii.mdr@gmail.com

ABSTRACT

This paper describes the prototype implementation of a pervasive, wearable augmented reality (AR) system based on a full body-motion-capture system using low-power wireless sensors. The system uses body motion to visualize and interact with virtual objects populating AR settings. Body motion is used to implement a whole body gesture-driven interface to manipulate the virtual objects. Gestures are mapped to correspondent behaviors for virtual objects, such as controlling the playback and volume of virtual audio players or displaying a virtual object's metadata.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Artificial, Augmented and Virtual Realities

H.5.2 [**User Interfaces**]: Input Devices and Strategies

General Terms

Algorithms, Design, Experimentation.

Keywords

HCI, User Interaction, Orientation, Augmented Reality, Mobile Computing, Body Area Networks, Wireless Sensor Networks, Upper Body Gesture Recognition

1. INTRODUCTION

There is significant interest in the development of more "natural" methods for Human Computer Interaction. Keyboards, joysticks, mice, displays and other devices are hardware interfaces in widespread use. Many intelligent technologies like speech and image recognition systems are commercially available to facilitate interaction through the use of naturalistic human-computer interface modalities. One interaction modality that been the focus of considerable research lately is that of Gestural Interaction, where commands from mouse and keyboard might be replaced with a user's gestures [2].

Virtual reality (VR) has been a focus of research and commercial practice for a number of years, not only for entertainment purposes but also for industrially relevant applications such as 3D product design and visualization. Recently, the approach of Augmented Reality has emerged in the computer science world as an extension to VR, where virtual worlds and objects, or worlds and metadata are mapped on to views of the real world, mixing the real with the artificial. However, both types of visualization suffer from problems of control – how can a user manipulate virtual objects as naturally as they would manipulate real physical ones? We aimed to examine the concept of naturalistic interaction with virtual objects in an AR setting by investigating how a wireless- sensor-based system could be used to recognize gestures made by the user's body and help create a wearable AR system that could be deployed and used without the need for fixed infrastructure.

The approach we took was to develop a system based on wireless sourceless sensor packs attached to the user's body to trap body motion, a VR headset and a web camera attached to the user's head. The sensors are provide raw data subsequently used for the recognition of the user's gestures, whilst the camera gives a live video feed on which virtual objects are superimposed. The web camera works with the sensor on the user's head to obtain the camera's orientation and as such, synchronize the panning and rotation of the virtual world to match the web camera movements. Our system is comparable to existing commercial offerings (e.g. XSens, EoBodyHF). These systems use sets of wired sensor packs, connected to a wireless hub, which transmits aggregated data wirelessly using Bluetooth or 802.15.4 respectively. Our system's advantage is that all sensors are wirelessly connected to a coordinator/transmitter node, which allows for improved wearability and flexibility in the configuration of the system, for full or partial body motion capture. Our motion capture system is described more extensively in [5]. In this paper we present the application of our system into combining the real world with a virtual one, as an enabler for Augmented Reality applications. We also present an outline of an application where this has been successfully trialled and discuss future opportunities.

2. BACKGROUND

AR technology enhances a user's perception of and interaction with the real world. The key feature of the augmented reality technology is to present auxiliary information in the field of view

for an individual automatically, without human intervention. The virtual objects display information that the user cannot directly detect with his or her own senses. The information conveyed by the virtual objects helps the user to perform real-world tasks. This new form of human-computer interaction can be applied to various industry areas [1]. AR is an emerging technology area and as such, applications that could benefit from AR technology are still not fully explored. Typical examples are seen in engineering, military and educational domains. AR technology for digital composition of animation with real scenes is being explored to deliver adverting content[1] or bring new digital entertainment experience to users.

Due to the recent standardization, miniaturization, modularisation and economies of scale presented by the new technologies available for the creation of wireless sensor networks, modern silicon Micro-Electro-Mechanical Systems (MEMS) sensors (such as gyros, magnetometers and accelerometers) all have become inexpensive, small and now can be worn on the body or integrated into clothing [12]. This represents exciting opportunities for HCI researchers engaged in interaction design research. For the purposes of AR, a sensor, coupled with a web camera, provides evident opportunity for orientation in a virtual world accordingly to the direction that camera faces. A skeletal model built from the sensors' data supplies the receiver with rotation matrices and linear angles that can be used to recognize human gestures [3].

As a part of our research group's current interest areas (pervasive healthcare, mobile spatial interaction and mobile audio-video interaction) our work aims to investigate the use of a low-cost distributed computing infrastructure with sensors to provide a means of wearable and pervasive human body motion capture with a system that we called Mirrored Motion (MM) [5]. Our developed system has been used in applications like gait analysis and body motion capture. Based on the work carried out to develop MM, we aimed to extend the surrounding environment with supplementary information through AR. Additionally, we aimed to use the same system not only to help visualize virtual objects for AR, but also interact with them through gestures.

3. SYSTEM DESIGN

AR technology is not a new concept. Apart from studies in AR visualization, many applications already exist in advertising, industry, robotics, medicine, sport, military and many other spheres. Additionally, several researchers have proposed to use gesture recognition in conjunction with AR [11] [12]. However, we are not yet aware of any AR systems based on full body motion capture and that utilize gesture interaction, which do not require extensive infrastructure support and which can be used in pervasive computing settings. From reviewing the existing literature, we identified defined two goals [7] [8] to be implemented:

- **Gesture recognition.** The proposed system must recognize user's gesture in 3D space independently on the user's location. Gestures can be trained before recognition.

- **Extending reality.** The system must provide means for presenting auxiliary information in the field of view of a user's VR headset. Providing the particular gesture is recognized the system is to change the state of correspondent virtual object.

---

[1] BMW Z4 advertisement: www.bmw.co.uk/z43d (link valid 5/09)

In designing the AR demonstration, we considered these goals as appropriate to inform our system characteristics.

3.1 Gesture recognition

Our system is comprised of sensor "nodes" that can be attached to key locations on a user's body, monitoring the movement of major body parts (limbs, torso, and head). One of the off-body nodes acts as a "coordinator", gathering data from all nodes and relaying to external systems (e.g. a PDA, a server or a desktop computer) for further processing. The approximate frequency of streaming data is 20 Hertz. While our system is capable of full body motion capture, in this work we used an upper body set of sensors, as we were more interested in torso and hands gesture recognition. An internal processing system provides us with an updatable skeleton model of the user [5], which is a method also used by other researchers, e.g. [6].

The posture of the skeleton is calculated in real-time through forward kinematics. Kinematics simplifies computations by decomposing any geometric calculations into rotation and translation transforms. Orientation is obtained by combining (or fusing) these information sources into a rotation matrix – an algebraic format that can be directly applied to find the posture of the user. The result is a simple skeletal model defined as a coarse representation of the user.

In general terms, gesture recognition consists of several stages, like feature extraction, preprocessing, analyzing and making a decision. Our experimental method consists of using linear angles between any two links in the skeletal model as a dataset that is fed into the gesture recognition algorithms described below.
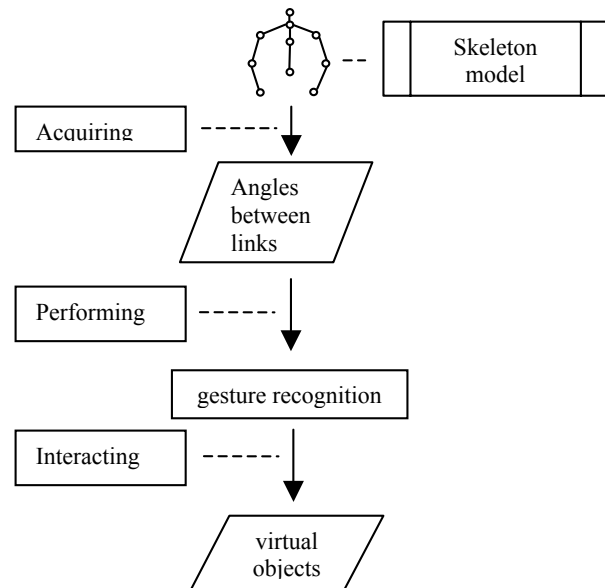

Figure 1. Gesture recognition architecture.

At the preprocessing stage we perform work to filter the noise caused by rapid movements and inaccuracy of the measurements (around 3-5 degrees). A magnetometer-accelerometer-gyro sensor can produce accurate orientation information when the forces experienced by the sensor are gravity or low accelerated movements. Any additional forces will result in the reference vector produced by the accelerometer to be inaccurate.

Analyzing sequences of linear angles and performing the gesture recognition itself was implemented with the help of AMELIA general pattern recognition library [13], which we used as a basis to implement our own customized Hidden Markov Model. Hidden

Markov models (HMMs) are the basis of most gesture recognition algorithms used today. However, traditional HMM-based gesture recognition systems require a large number of parameters to be trained in order to give satisfying recognition results. In particular, an **n**-state HMM requires n^2 parameters to be trained for the transition probability matrix, which limits its usability in environments where training data is limited [4][9]. The reduced model that was used in our system uses a constant number of parameters for each state to determine transition probabilities between all states. As there are many different notation conventions in use for Hidden Markov Models, here we utilize a convention we believe makes our model easy to understand. We thereby define our augmented hidden Markov model ($S = \{E,N\}$, S_b, S_e, T,O) by a set of states S, a designated beginning state S_b, a designated ending state S_e, a state transition probability function T , and an observation probability function O. The augmented HMM behaves essentially the same as a regular HMM, with only a few points of departure. The set of states S is divided into disjoint sets of emitting states E and non-emitting states N. The difference between the two is that when entered, emitting states emit an observation belonging to the observation set θ according to the observation probability function $O : E \times \theta \rightarrow [0,1)$. The model always begins in the beginning state $S_b \in S$, and until it ends up in the ending state $S_e \in S$ it makes transitions according to the transition probability function $T : (S - S_e) \times S \rightarrow [0,1)$. T must also satisfy that the sum of transition probabilities out of any state is 1. In the reduced parameter model, we use the following parameters, depicted also in Figure 2:

- τ_i is the probability of remaining in an emitting state, $T(E_i, E_i)$
- η_i is the probability of going to the next emitting state $T(E_i, E_i + 1)$
- ς_i is the probability of skipping at least one emitting state $T (E_i, N_i+1)$
- k_i is the probability of skipping an additional emitting state $T (N_i, N_i+1)$
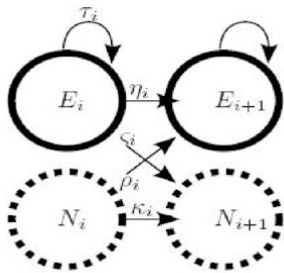- p_i is the probability of ending a skip sequence $T (N, E_i+1)$

Figure 2. Transition probability parameters for the HMM model.

Our system allows users to record their own gestures for predefined actions that control the behaviour of virtual objects (e.g. selecting/deselecting an object, turning on and off a virtual appliance such as a virtual radio, controlling the behaviour of a virtual object such as start/stop playback of music), some of which are depicted in Figure 3. As such, different actors may use diverse gestures for the same action. Typically, to record one gesture an actor repeats it for 3-4 times, as in [4] [10]. Once a few "recordings" of a gesture have been made, the system is then trained on the captured motion data set in order to be able to recognize the gestures. A general gesture tends to be 2 to 3 seconds in time. After training, the user can perform gestures in

different sequences as well as performing actions that are not gestures. Our system recognizes gestures with the probability of 80-90% (determined experimentally). Examples of our gesture recognition systems are available to view online in video form[2].
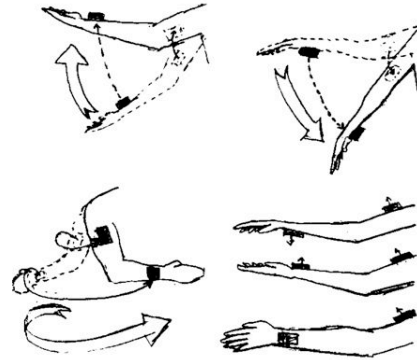

Figure 3. Examples of naturalistic gestures designed to control a virtual radio in the AR system. The top gestures show raising the (left) and lowering (right) the volume. The bottom left shows skipping a station. The bottom right shows that by modifying the position of just one node (carpal), we can achieve a large number of distinct gestures depending on how that node moves in 3D space.

At this point in time, our system has two limitations: Firstly, saving of the recorded gestures training data is not yet implemented (due to development-time constraints) but we consider it as a simple goal. Secondly, our current recognition model does not allow a gesture to stop in the actor's relaxed position. For example, if a user stands still and tries to record a gesture, finishing it at the relaxed posture, the recognition system will not determine when the gesture ends. However, this limitation will be removed in the near future.

3.2 Extending reality

In order to provide a synthesis of live video feed and virtual objects within the system, an XNA gaming platform was used. Live video that comes from a web camera is constantly placed in front of a viewer in a 3D world. In order to ensure that the 3D world's game camera corresponds with some fidelity to the live video feed from the webcam, the system must be calibrated by starting at a pre-determined real-world location whose coordinates are mapped to a pre-determined point in the virtual world. The user sees a combined image from real video and virtual objects. Virtual objects are placed in front of the dynamic web camera feed. The coordinates of the video are not updated, therefore live view always stays on the same place - in front of the viewer, whereas coordinates of the virtual objects are updated. We combined a web camera with the head sensor, which helps map the camera orientation (and hence the user's view of the real world) in 3D space. As a user moves his or her head, the virtual world moves accordingly. The virtual objects that are in front of the human actor will come in and out of the user's field of view, when the viewer looks to the left or to the right (up or down).

Currently, to select a virtual object, we used data that comes from the sensor on the right hand. We transform pitch and rotation to the Y and X movements of a cursor in the virtual world. To select a virtual object, the user thus uses their right arm to "point" the cursor to the virtual object they want to select. Every virtual object has its own bounding form. For simplicity, we used

bounding spheres only. We took advantages of the XNA ray technique to understand whether a ray (game camera – cursor – infinity) intersects with the bounding spheres of virtual objects. When the cursor line of sight intersects and hovers over an object, it becomes selected (Figure 4).

Figure 4. The user points an arm-controlled cursor (visualized to the user as an X) at the virtual object (marked by the white arrow), which is then highlighted. Metadata for that object is subsequently displayed.

4. CONCLUSIONS & FURTHER WORK

In this paper we described how we implemented gesture recognition with a pervasive body motion capture system and created augmented reality, which might be used in different fields such as entertainment, sports, military etc. Throughout our development we aimed to make use of easily available, low-cost components, keeping the cost per node to approximately £150. Our proposed system shows the way we wished people to interact with and benefit from our work. Overall we were successful in delivering a high-performance, truly pervasive, extensible and highly wearable system that fulfils the criteria for augmented reality systems. In Figure 4, the user's only restriction to mobility is the headset connection, in this picture connected to a desktop PC, but equally able to be connected to a portable laptop/tablet. However, our system at the moment does not support the motion of the user's body between locations in the real/virtual world. We assume that the user remains fixed and as such we have only used the upper body sensor set as a means to trap gestures. In the near future, we plan to take advantage of our ability to capture motion from the entire body, in order to allow the user to move through the AR world. We would be particularly interested in examining how our MEMS-based system performs in inferring user location (e.g. while walking) and how the accuracy of our system might be enhanced through the fusion of GPS data, where available. Additionally, a hybrid positioning system as described would be of great interest to examine in scenarios where indoor-outdoor transitions occur for the user.

We believe that our system will prove an extremely useful tool for a range of interaction opportunities; aside from our previous projects we are working on applying our system in several areas. We are particularly interested in its potential in mixed reality situations for gaming, We also wish to investigate issues in human-human interaction through embodied agents, controlled through the motion capture system. We are looking into the control of VR agents, as well as robotic agents for which the metaphor of "transferring one's soul" will be used to investigate response and interaction with other humans. Finally, we are interested in pursuing applications in tangible interfaces and semi-virtual artifacts, as well as gesture-based whole-body interaction with large situated displays. We hope to be able to create new types of human-computer interfaces for manipulating program windows, arranging or opening files using ad-hoc large projected or semi-transparent situated displays.

5. REFERENCES

[1] Lyu M.R., King I., Wong T.T., Yau E., Chan P.W., 2005. ARCADE: Augmented Reality Computing Arena for Digital Entertainment. 5th IEEE Aerospace Conference, Big Sky, MT, USA

[2] R. Azuma, "A survey of augmented reality" Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp. 355-385, August 1997.

[3] S. Rajko, G. Qian, HMM Parameter Reduction for Practical Gesture Recognition, IEEE International Conference on Face and Gesture Recognition, 2008.

[4] S. Rajko, G. Qian, T. Ingalls and J. James, Real-time Gesture Recognition with Minimal Training Requirements and On-line Learning, IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[5] Philip Smit, Peter Barrie, Andreas Komninos. Mirrored Motion: Pervasive Body Motion Capture using Wireless Sensors. Whole Body Interaction Workshop, ACM CHI2009, Boston MA.

[6] Crossan, A., Williamson, J., Brewster, S., Murray-Smith, R., 2008. Wrist rotation for interaction in mobile contexts. In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services. Amsterdam, The Netherlands: ACM, pp. 435-438.

[7] Antifakos, S. & Schiele, B., 2002. Bridging the gap between virtual and physical games using wearable sensors. In Proceedings of the Sixth International Symposium on Wearable Computers, 2002. (ISWC 2002), p. 139-140.

[8] Seon-Woo Lee & Mase, K., 2002. Activity and location recognition using wearable sensors. *Pervasive Computing, IEEE*, 1(3), p.24-32.

[9] Molet, T., Boulic, R., Thalmann, D., 1999. Human Motion Capture Driven by Orientation Measurements. Presence: Teleoperators and Virtual Environments Vol.8(2), pp.187-203.

[10] Bodenheimer, B., Rose, C., Pella, J., Rosenthal, S. 1997. The process of motion capture: Dealing with the data. Computer Animation and Simulation. pp. 3-18

[11] Tognetti, A., Lorussi, F., Tesconi, M., Bartalesi, R., Zupone, G., De Rossi, D., 2005. Wearable kinesthetic systems for capturing and classifying body posture and gesture. Conf Proc IEEE Eng Med Biol Soc. 2005 ;1 :1012-5.

[12] Buchmann, V., Violich, S., Billinghurst, M., Cockburn, A. 2004. FingARtips: gesture based direct manipulation in Augmented Reality. In: GRAPHITE 2004: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia.

[13] AMELIA: A generic library for pattern recognition and generation: http://ame4.hc.asu.edu/amelia/ (link valid 5/09)