Markerless Full Body Tracking: Depth-Sensing Technology within Virtual Environments

Belinda Lange, Skip Rizzo, Chien-Yen Chang, Evan A. Suma, Mark Bolas
Institute for Creative Technologies
Playa Vista, CA
lange@ict.usc.edu, arizzo@usc.edu, kchang@ict.usc.edu, suma@ict.usc.edu, bolas@ict.usc.edu

ABSTRACT

Over the last decade there has been growing recognition of the potential value of virtual reality and game technology for creating a new generation of tools for advancing rehabilitation, training and exercise activities. However, until recently the only way people could interact with digital games and virtual reality simulations, was by using relatively constrained gamepad, joystick and keyboard interface devices. Thus, rather than promoting physical activity, these modes of interaction encourage a more sedentary approach to playing games, typically while seated on the couch or in front of a desk. More complex and expensive motion tracking systems enable immersive interactions but are only available at restricted locations and are not readily available in the home setting. Recent advances in video game technology have fueled a proliferation of low-cost devices that can sense the user’s motion. This paper will present and discuss three potential applications of the new depth-sensing camera technology from PrimeSense and Microsoft Kinect. The paper will outline the technology underlying the sensor, the development of our open source middleware allowing developers to make applications, and provide examples of applications that enhance interaction within virtual environments and game-based training/rehabilitation tools. The PrimeSense or Kinect sensors, along with open source middleware, provide markerless full-body tracking on a conventional PC using a single plug and play USB sensor. This technology provides a fully articulated skeleton that digitizes the user’s body pose and directly quantizes their movements in real time without encumbering the user with tracking devices or markers. We have explored the integration of the depth sensing technology and middleware within three applications: 1) virtual environments, 2) gesture controlled PC games, 3) a game developed to target specific movements for rehabilitation. The benefits of implementing this technology in these three areas demonstrate the potential to provide needed applications for modern-day warfighters.

ABOUT THE AUTHORS

Belinda Lange, PhD is a Senior Research Associate in the Medical Virtual Reality Group at the Institute for Creative Technologies and holds a Research Assistant Professor position in the School of Gerontology at the University of Southern California. She received her PhD and degree in Physiotherapy (with Honors) from the University of South Australia and her Science Degree from Flinders University of South Australia.

Albert ‘Skip’ Rizzo, PhD directs the Medical Virtual Reality Group at the University of Southern California’s Institute for Creative Technologies. He is Research Professor - Psychiatry/Gerontology. Dr. Rizzo conducts research on the design, development and evaluation of Virtual Reality systems targeting the assessment and training/rehabilitation of spatial abilities, attention, memory, executive function and motor abilities.

Chien-Yen Chang is a Lead Engineer in the medical Virtual Reality Group at the Institute for Creative Technologies. His work spans from fully immersive virtual environments to integration of low cost devices within active video games.

Evan A. Suma, PhD is a postdoctoral researcher in the areas of human-computer interaction, cognition, and perception in virtual environments and 3D user interfaces. Evan graduated from Ithaca College with a BA in computer science and received his PhD from the University of North Carolina at Charlotte.

Mark Bolas is Associate Professor of Interactive Media in the School of Cinematic Arts, Associate Director of the Institute of Creative Technologies at the University of Southern California, and is Chairman of Fakespace Labs, Inc. Mark majored in Physics and minored in Music at University of California, San Diego and holds a terminal MS in Mechanical Engineering (design) from Stanford University.
Markerless Full Body Tracking: Depth-Sensing Technology within Virtual Environments

Belinda Lange, Skip Rizzo, Chien-Yen Chang, Evan A. Suma, Mark Bolas
Institute for Creative Technologies
Playa Vista, CA
lange@ict.usc.edu, arizzo@usc.edu, kchang@ict.usc.edu, suma@ict.usc.edu, bolas@ict.usc.edu

INTRODUCTION

Over the last decade there has been growing recognition of the potential value of virtual reality (VR) and video game technology for creating a new generation of tools for advancing rehabilitation, training and exercise activities. Part of the excitement in this area has derived from the idea that by providing people with the opportunity to practice cognitive and motor exercises within a digital game simulation, they would be more motivated to regularly engage in such recommended activities. However, until recently the only way people could interact with digital games and virtual reality simulations, was by using relatively constrained gamepad, joystick and keyboard interface devices or complex and expensive motion capture systems. Thus, rather than promoting healthy physical activity in accessible locations, these modes of interaction served to encourage a more sedentary approach to playing games, typically while seated on the couch or in front of a desk. Therefore, while people were drawn to the fun and excitement offered by the rapid evolution of modern computer games capable of delivering digital experiences that one could only dream about back in the bygone days of the 20th Century, such games didn’t provide much in the way of immersive physical interactions.

Recent advances in video game technology have fueled a proliferation of low-cost devices that can sense the user’s motion. These range in capability from handheld controllers that can be used for gesture-based control, such as the Nintendo Wii-mote and the PlayStation Move, to cameras that use computer vision techniques to sense the user’s body pose in a two-dimensional plane, such as the Playstation® EyeToy™. In the past year, low-cost depth-sensing cameras have also become commercially available, including the widely publicized Microsoft Kinect, which have made it possible to sense the full-body pose for multiple users without the use of markers or handheld devices. OpenNI and Microsoft Windows have made available an open source framework for developers to access and develop upon the technology. To facilitate the rapid development of virtual reality applications using OpenNI-compliant devices (currently the PrimeSensor, Microsoft Kinect and soon the Asus Wavi XTion), as well as to incorporate motion-based control in existing off-the-shelf games, we have developed the Flexible Action and Articulated Skeleton Toolkit (FAAST). This toolkit will be available as an open source software and it should be noted that the authors of this paper are not selling or marketing the hardware or software described within this paper.

CURRENT ISSUES IN REHABILITATION

Optimal rehabilitation therapy involves a high dose of skilled repetitive exercise. High demands are placed on the clinician’s time, professional knowledge and experience to find and work out the most effective and appropriate methods of treatment for the individual patient. Resources are limited and it can be difficult for clinicians to provide these exercises during one on one time in the clinic setting. Repetitive skilled practice is crucial for rehabilitation, often requiring patients to perform exercise programs as ‘homework’ outside of standard therapy sessions (Weightman et al., 2010). Home-based or clinic-based exercise programs are often provided to patients in the form of a list of exercises with pictures and instructions. However, the patient is responsible for completing the exercises accurately and reporting progress back to the clinician. Adherence to self-guided exercise programs in the clinic and home settings is notoriously low and very difficult to quantify due to the reliance on patient’s subjective feedback and reliance on patients maintaining accurate recording of their exercise sessions in an exercise diary.

Video Games and Virtual Reality

The use of video games and virtual reality systems for rehabilitation has expanded rapidly over the past five years. Early research in the area of virtual reality
to assist people to relearn how to move, suggests that virtual reality game-based technology can be used effectively to improve motor skill rehabilitation of a range of functional deficits (Boian et al., 2002; Chuang et al., 2002; Adamovich et al., 2004; Dvorkin et al., 2006; Fung et al., 2004; Fulk et al., 2006; Fung et al., 2006). Virtual reality systems demand focus and attention, can motivate the user to move, and provide the user with a sense of achievement, even if they cannot perform that task in the 'real world'. Current rehabilitation practices at medical centers (such as the Walter Reed Army Medical Center, National Intrepid Center of Excellence, BAMC Center for the Intrepid and Naval Health Research Center) include training with therapeutic technologies that challenge survivors of blasts by exposing them to virtual reality stimuli using training devices such as the Computer-Assisted Rehabilitation Environment (CAREN) to develop and evaluate protocols to train balance, gait and dual task training programs for Service members with TBI (Scherer & Schubert 2005). While these high-end virtual reality technologies have the potential to contribute a wealth of knowledge about the development and implementation of game-based training programs, the ability for patients and clinicians to access the system outside of clinic hours and once the inpatient treatment program has been completed is limited. However, video game technology is improving in a way that allows the development of low-cost game-based applications that can complement existing high-end virtual reality applications like the CAREN system and increase access within the everyday clinic and home settings. Clinic and home-based systems need to be affordable and easy to set-up, use and maintain, while still providing the most appropriate and accurate interactions so that the user can practice motor activities needed to foster transfer to the real world. The recent release and worldwide acceptance and enjoyment of Nintendo® Wii™, Nintendo® WiiFit™, Sony Playstation®2 EyeToy™, Sony Playstation®3 MOVE™ and Microsoft® Xbox360 Kinect™ has provided significant evidence for the notion that exercise can be fun, provided it is presented in a manner that is entertaining, motivating and distracting.

Exercise-based video games have the potential to increase availability and enjoyment of repetitive exercise for rehabilitation at a low cost. Off-the-shelf games for commercial gaming consoles have been developed and tested for the purpose of entertainment; however, the games and consoles were neither designed as medical devices nor with a primary focus of an adjunct rehabilitation tool. While games on these consoles were not designed with rehabilitation in mind, they have the advantage that they are affordable, accessible and can be used within the home. Furthermore, by using these devices for exercise, individuals have anecdotally reported a high level of enjoyment by interacting and exercising with friends and family members. Some researchers have treated neurological impairments by implementing off the shelf game consoles, such as the Sony Playstation®2 EyeToy™ (Flynn et al., 2007; Haik et al., 2006; Rand et al., 2008) and Nintendo Wii (Deutsch et al., 2008), with promising results. The underlying motion-sensing and 3D graphic technologies that are used in these commercial game systems allow the user to engage in entertaining motor games using gross body movements that are not bound by the limits of a mouse, joystick or game-pad interface. Yet whilst these systems have enjoyed wide adoption by millions of users and are generally stimulating (Graves et al., 2008) and entertaining, clinicians and patients cannot easily alter the hard coded stimulus parameters of the game system as is needed to optimally rehabilitate precise motor skills beyond simple flailing. In addition to the limited options for the systematic control of stimulus parameters needed to customize interaction challenges to the needs of the user, they provide limited capacity for the recording of meaningful performance data (Lange et al., 2009). So while these interaction games are fun and motivating, such ‘out-of-the-box’ systems do not meet the theory-driven requirements for delivering precise motor interventions in a systematic fashion that can be titrated to the needs of a target user group. However, the potential does exist that these systems can be creatively repurposed for useful rehabilitation purposes

**Game-Based Rehabilitation**

Within the Medical Virtual Reality Group Game-Based Rehabilitation Lab and MxR Lab at the Institute for Creative Technologies, we have been addressing the challenge for creating low-cost home-based video game systems for motor assessment and rehabilitation. The focus of this research is three-fold: 1) Assess the usability of off-the-shelf games and consoles within a range of user populations, particularly people with Traumatic Brain Injury, Spinal Cord Injury and Amputation; 2) Using this feedback, re-purpose or develop low-cost interaction devices that are appropriate for use within the rehabilitation setting; 3) Design, develop and test games specifically focused on rehabilitation tasks. The key advantage of designing these games is to provide the therapist and/or patient with the ability to alter elements of game play in order to tailor treatment tasks for individual users and expand the use of these tasks to a wider range of level of ability. The ability to store data and provide information
about performance is another feature that supports the development of specific games for rehabilitation. The use of games for rehabilitation must maintain the goals of existing therapies, whilst improving motivation to perform therapeutic exercise programs.

The most recent application we have been developing leverages the technology of the PrimeSense 3D depth-sensing camera (Lange et al., 2011). We have developed a flexible framework that allows for integration of the 3D depth-sensing technology of the PrimeSense camera or the Microsoft Kinect sensor with associated software to drive any PC-based computer game through tracking of the full-body skeleton without any markers. This allows the user to control a virtual character on the screen that directly represents their movements in the real world. The application requires only the 3D depth-sensing camera and USB connection to a PC.

DEPTH-SENSING TECHNOLOGY WITHIN VIRTUAL ENVIRONMENTS

Hardware
Full-body interaction within virtual reality and game-based environments can be provided by the PrimeSense Reference Design (Figure 1) or the Microsoft Kinect sensor (Figure 1), USB plug-and-play devices that use an IR projector along with standard RGB and infrared CMOS image sensors.

To construct a depth map, the sensor uses a proprietary algorithm to resolve the pattern produced by projecting coded infrared light onto the scene geometry. This system has a field-of-view of 58 degrees horizontal and 45 degrees vertical, and generates depth maps with a resolution of 640x480 at 30 frames per second.

Software Architecture

The research team has been developing a Flexible Action and Articulated Skeleton Toolkit (FAAST) with the goal to make a general purpose software environment that enables many applications to quickly be modified to use depth-sensing technologies (http://projects.ict.usc.edu/mxr/faast/). The FAAST is middleware to facilitate integration of full-body control with virtual reality applications and video games using OpenNI-compliant depth sensors (e.g. PrimeSensor, Microsoft Kinect and Asus Wavi XTion). FAAST incorporates a VRPN server for streaming the user's skeleton joints over a network, which provides a convenient interface for custom virtual reality applications and games. This body pose information can be used for goals such as realistically puppeting a virtual avatar or controlling an on-screen mouse cursor (Figure 2).

Additionally, the toolkit also provides a configurable input emulator that detects human actions and binds them to virtual mouse and keyboard commands, which are sent to the actively selected window. Thus, FAAST can enable natural interaction for existing off-the-shelf video games that were not explicitly developed to support input from motion sensors. The actions and input bindings are configurable at run-time, allowing the user to customize the controls and sensitivity to adjust for individual body types and preferences. In the future, we plan to substantially expand FAAST's action lexicon, provide support for recording and training custom gestures, and incorporate real-time head tracking using computer vision techniques.

FAAST was initially developed to provide a convenient and accessible interface for the PrimeSensor Reference Design, a USB plug-and-play depth-sensing camera developed by PrimeSense. This technology, based on infrared structured light to compute a depth image of the environment, was licensed to Microsoft for the Kinect. Initially, FAAST relied exclusively upon OpenNI software, along with NITE middleware provided by PrimeSense, to perform user identification, feature detection, and basic gesture recognition using the depth image from the sensor. FAAST interfaces directly with OpenNI/NITE to access this information and performs additional high-level gesture recognition for generating events based on the user's actions. However, the recent release of the Microsoft Research Kinect SDK Beta on Windows platforms has introduced an alternative skeleton tracking implementation.
specifically for the Microsoft Kinect sensor. As each of these software libraries have their own advantages and disadvantages, FAAST now provides the capability for the user to select which skeleton tracking implementation to employ.

FAAST considers two broad categories of information from the sensor: actions and articulated skeletons. Articulated skeletons consist of the positions and orientations for each joint in a human figure and are useful for virtual reality and video game applications in allowing direct body-based control of a virtual avatar. FAAST retrieves these skeleton joints from the OpenNI drivers and transmits them to the end-user application using the Virtual Reality Peripheral Network (VRPN), a popular software package in the virtual reality community for interfacing with motion tracking hardware (Taylor et al., 2001). We built a custom VRPN server into the FAAST application that streams the skeletal information for each joint as a six degree-of-freedom tracker, allowing applications to interface with the sensor as they would any other motion tracking device. Figure 3 shows an example user puppeting a virtual wireframe avatar and a skinned virtual character rendered using an existing virtual reality engine with a VRPN client. To interface with existing applications and games, FAAST also can control the Windows mouse cursor by making it follow the user's hand joints with either absolute 1:1 motion over a customizable "active area" or relative motion with velocity determined by hand position.

![Figure 3. User puppeting a virtual character](image)

In contrast to articulated skeleton data, actions are more complicated since they require inferring meaning from the user's pose and their movements over time. NITE provides some basic gesture recognition for certain atomic actions involving the user's hands, such as a push, swipe, circle, and wave. These actions appear well suited for simple 2D interfaces such as media center control, but 3D selection, manipulation, and locomotion requires a much richer gesture set. Thus, we calculate several other relevant actions based on the user's body pose, including lean (forward, backwards, left, and right) and twist (left and right). Within FAAST, the user can adjust the angular threshold for detecting these actions to make them more or less sensitive. Actions were also developed for arm or leg movements in different directions, jumping, and walking-in-place. Any of the actions computed in FAAST, as well as the basic actions from NITE, can be bound to virtual keyboard or mouse commands that are sent to the actively selected window. Thus, FAAST enables these custom sensors to provide input to arbitrary applications, including off-the-shelf games that were not developed to support motion sensing input. Furthermore, users can customize the bindings and sensitivity for these actions at run-time, thus providing flexible input that can easily be adjusted according to the individual user's body type and preferences.

Sample Applications

The FAAST system can be used to emulate standard mouse and keyboard commands, triggered by a designated physical activity (i.e., body posture and specific gestures) of the user. This allows the user to add custom body-based control mechanisms to existing off-the-shelf games that do not provide official support for depth sensors. Figures 4 and 5 demonstrate users playing standard games (Tux Racer and Space Invaders) using body gestures instead a keyboard or mouse.

![Figure 4. User controlling a standard game with gestures through keyboard emulation option in FAAST](image)

![Figure 5. User controlling a standard game with gestures through keyboard emulation option in FAAST](image)
Figure 6 demonstrates an early application using the Microsoft Kinect/PrimeSense sensor as a tool to capture naturalistic user activity for interaction with the popular World of Warcraft (WoW) game. The user can assign any specific body gesture or movement as an action to interact in this game. Essentially the software allows a gesture to be assigned to emulate the keyboard/mouse actions that are normally used for interaction with WoW. While the complexity of all possible WoW keyboard/mouse actions may be challenging to integrate into quickly learnable body actions, users have reported thus far that even with limited coverage, the embodied interaction with the game was compelling. Keep in mind that this is part of a larger research program on 3D User Interaction and that we don’t have any illusions that hardcore gamers are going to toss their keyboards and other interaction devices in mass to play WoW in this manner. But, the ability to naturally interact with digital content using full-body action presents considerable opportunities for creating applications that address useful purposes beyond just entertainment. For example, such systems can be implemented in virtual rehabilitation for motor impairments. Following a stroke, brain injury or other form of neurological disorder, a person can naturally interact with game content as part of their physical, occupational or cognitive therapy, allowing them to practice complex movement integration within an online virtual environment such as WoW or Second Life.

The game begins with a calibration step, which is important in order to tailor the game to individual players’ level of ability. This step is guided by a clinician to encourage movements that are appropriate for the player and specific therapy goals. In the game, all eight jewels are present and visible in a series of rings that extend as the player moves through the mineshaft. Individual jewels glow to indicate when they can be gathered (Figure 8). The order in which the jewels glow is controlled by three different pre-defined patterns that the clinician can select before a session begins.

Before starting the game the clinician is given the option of displaying the player’s avatar as either a full torso with arms and head or simply displaying the hands (Figures 9a and 9b).
This avatar representation allows the clinician to control how much feedback the player receives about their body position during the game. The full upper torso is tracked within both representation settings, however the hands only representation provides visual representation of the hand points of the skeleton, providing limited external feedback to the player, encouraging internal focus of attention to the control of the movement, a goal for progressing existing static weight shift and balance training activities in the clinic. Additionally, the duration of time it takes patients to complete each jewel ring task and the total time it takes to complete the game are recorded and reported (Figure 10).

**Figure 10. Data recorded from game**

**Initial User Feedback**

To date, 23 patients (19 male, 4 female) with balance issues related to Stroke (n = 10), TBI (n = 4) and SCI (n = 9) and 10 clinicians (4 male, 6 female) have provided feedback on the game-based prototypes. Overall, patients reported the games to be challenging and fun, stating they would be likely to use the technology within the clinic and home settings if given the option. Feedback was provided about the game, including issues and potential solutions (Lange et al., 2011). Clinicians stated they were excited about the use of this type of technology within the clinical setting. The option to assign specific gestures or calibrate the system for different individuals, vary the task within the game and change the avatar view and feedback appealed to the clinicians.

**CONCLUSION**

Video game technologies may provide a means to create rehabilitation and exercise environments that will allow health care professionals to precisely deliver and control complex dynamic, interactive, 3D stimulus presentations. For rehabilitation, the simulated experience may be designed to address certain aspects of sensory or motor loss experienced by an individual. Video game technologies can record targeted responses permitting assessment of functional performance. Effective game-based rehabilitation that is easily adapted for individuals to use in inpatient, outpatient and home-based care could be a valuable adjunct to conventional therapy. We believe that the integration of game-based rehabilitation with low cost sensing systems for tracking human movement will revolutionize how motor rehabilitation and exercise activities will be done in the future by allowing more therapeutic exercise to occur in the home. This technology has the potential to provide a new dimension for interactive exergaming that could be useful as one aspect of a multicomponent approach for reducing adult and childhood obesity and subsequently reducing the incidence of diabetes that might be preventable with regular exercise and weight control.

The use of low-cost game-based rehabilitation tools such as those outlined in this paper demonstrate the potential for integration of the use of these systems for rehabilitation of service members following discharge from inpatient care. This technology could also be used to develop tools to guide and motivate exercise activities for service members, providing options for training during periods of extreme weather. The technology demonstrated in this paper could also be used as a low-cost training tool for a range of tactical individual and team-based activities, allowing practice to take place outside traditional training locations.

**ACKNOWLEDGEMENTS**

The authors would like to thank the PrimeSense Corporation and OpenNI community. This work was supported by National Institute on Disability and Rehabilitation Research (NIDRR) grant H133E080024 (OPTT RERC, University of Southern California).

**REFERENCES**


