A Virtual Whole Body System

Bernhard Spanlang
EVENT Lab
Universitat de Barcelona

David Corominas
EVENT Lab
Universitat de Barcelona

Mel Slater
ICREA - EVENT Lab
Universitat de Barcelona

Figure 1 Left: a participant in motion capture suit, wearing a HMD; Middle:Avatar representation of the participant from third person view with mirror reflections and dynamic shadows; Right: First person view of avatar representation, with the avatar reflected in a mirror.

ABSTRACT
This paper is a practice and experience paper on the use of a low cost optical whole body motion capture system in combination with an ultrasound head tracker and a wide field of view head mounted display (HMD) that enables the user to visually replace his real body with that of a virtual humanoid or alien body. We describe the integration of the low cost Optitrack motion capture system with HALCA a hardware accelerated library for realistic character visualization and real time animation in a Virtual Reality (VR) system. We demonstrate our system by mapping in real-time the tracked whole body motions of a person onto a virtual character which is visualized dynamically with shadows and reflections in a HMD from a first person view. The system is used for neuroscience experiments on virtual body ownership.

Categories and Subject Descriptors
H.5.1. Multimedia Information Systems: Artificial, augmented, and virtual realities.

General Terms
Virtual body owndership, Human Factors.

Keywords
Motion tracking, virtual humans, body ownership, virtual reality.

1. INTRODUCTION
Since the early days of virtual reality that employed head-tracked head-mounted displays (HMD) the importance of a self-representation, an egocentric virtual body has been recognized – in particular with respect to participant's sense of presence within the virtual environment [15, 16]. This has also come to be recognized more recently with a new wave of interest in HMDs, for example [9]. A virtual body representation has many applications not only for VR and computer games but even in health care. However, in more common human-computer interaction, such as for computer games (e.g., Quake) it is normal that no virtual body is required because the interactive images are shown from a third person perspective, and the user does not have head tracking so that it is impossible to look down at one's own body - although the player may see a virtual body's (or avatar's) weapon or hands etc.

In virtual worlds such as Second Life[1] or World of Warcraft[2] the player usually can observe his avatar from a birds eye view and control is via the keyboard, a joystick or other simple interaction devices. More recent hardware for games such as the Nintendo Wii and the Wii balance board have shown new possibilities in controlling an avatar. Here, contrary to the usual hand controlled avatar the user of a Wii game can control avatar movement by moving the real body. Therefore, while limited, real physical exercise is executed when such games are played. Such systems use simplified tracking or in the case of the balance board a way of measuring the centre of gravity of the human participant by using four digital scales. Such information is then interpreted so that the avatar can mimic the movements of the player.

Recent neuroscience research indicates that the human brain can be tricked to believe that a static body part or a static whole body mannequin [8, 3] belongs to the human participant. First results on a whole body ownership illusion with a head tracked virtual character were shown in [14] The rubber hand illusion experiment was an early starting point of research in this direction [1]. Ehrsson et al [4] showed that particular brain areas are activated if a human participant is experiencing the rubber hand illusion.

This paper describes a system that we built in order to carry out further experiments on the impact of a dynamic virtual body on the sense of body ownership. The system consists of a low cost optical tracking system, a wide field of view HMD, whole body tracking software, a hardware accelerated library for character animation (HALCA) [13] and the XVR system [2].
Related work will be discussed in the next Section. The components of our system are then described in detail in Section 3. Results from people interacting with the system are given in Section 4. Finally conclusions and suggestions for future work are given in Section 5.

[1] http://secondlife.com/

[2] http://www.worldofwarcraft.com

2. PREVIOUS WORK

Two types of motion capture system have emerged over the last decade, marker based and more recently markerless systems. Marker-less systems can be categorized into skeletal capture systems[3] and 3D mesh capture systems[4].

Marker based and marker-less skeletal systems require a kinematic skeleton and a skinned body model that can be animated when visualization of a deformable body is required. Similar technology is used for whole body scanning[5] though the resulting meshes are usually static. Body scanning technology has been developed for well over a decade now and its main application at the moment is in the textile industry, for erognomics for example in the automotive and also in the entertainment industries.

Vlasic et al [17] recently presented a paper that describes a system that tries to merge marker-less skeleton tracking and mesh tracking functionality. Their approach however is not real time and requires more than 15 seconds of pure processing time per frame of animation. Horaud et al [5] demonstrate a system that can animate a deformable body model based on point cloud data extracted from a multi view camera system. They use a maximum likelihood model to estimate the pose of the deformable model.

A high accuracy very low cost full 6DOF tracking system that is based on the Wii –mote input devices of the Nintendo games console was presented by Hay et al in [6]. With current restrictions of the Wii mote this system allows to track only 4 markers, but at less than 100€ the system costs one to two orders of magnitude less than any other 6DOF tracking system. In this paper we use a low cost motion capture system that employs a very simple and fast calibration technique and which enables us to track the whole body of currently up to two participants.

3. SYSTEM COMPONENTS

Our system consists of four components, a low-cost full body motion capture system, an Intersense head tracking system, a hardware accelerated library for character animation (HALCA) and a wide field of view HMD. In the following subsections we detail these components.

3.1 Tracking Environment

The Optitrack full body motion capture system consists of multiple infrared cameras that are connected to a PC via USB. A minimum number of 6 cameras is recommended in order to track the 34 markers of the adaptable skeleton with predefined structure. In the next subsections we describe in more detail the properties of the environment required for the system to work and the cameras and software that we use.

3.1.1 Tracking Environment

The optical system cannot be used in every environment. The cameras are sensitive to infrared light which does occur in natural daylight. Therefore the system will not work in natural daylight or in rooms with transparent windows. Artificial light sources for example neon tubes do not radiate much infrared and therefore the system works well with such artificial light. However, the system is very sensitive to specularly reflecting surfaces. Usually the infrared source of one camera is visible by one or more other cameras. Since these sources are fixed they can be masked out by the camera software. However, infrared light rays that are reflected from a specular surface are difficult to mask and therefore have to be avoided as much as possible in the tracking environment.

3.1.2 Cameras

Optitrack V100 cameras consist of a 640×480 Black and White CMOS image capture chip, a interchangeable lens, an image processing unit, a synchronization input and output to synchronize with other cameras, an infrared LED source and a USB connection for controlling the camera capture properties and to transmit the processed data to a PC. Lenses with different FOV from 45-115degrees can be used with these cameras. Currently we have tested only with the default 45degrees lenses. The CMOS capture chip is sensitive to infrared light.

The cameras can deliver images at a frame rate of 100Hz. At this rate the USB can transmit the complete images of only one camera. In order to work with multiple cameras the images need to be pre-processed by a threshold method on the camera so that only 2D marker positions are transferred. Data delivered by say a 12 camera system is reduced from several Mb/s to a few Kb/s by this pre-processing step. Dedicated pre-processing also reduces the latency of the system. More traditional optical tracking systems can suffer from higher latencies if all images are transferred to a PC first and are processed there.

The Optitrack cameras can work in different image processing modes: object, segment and grayscale precision modes. Object mode is used to extract markers at high speed. Segment mode is for extracting reflective stripes, which is not common in motion capture. The precision mode delivers more grayscale information of the tracker position in order to be able in the software to identify the tracker position at higher precision.

3.1.3 Camera configuration

We use adjustable wall-camera-mounts that enable us to freely position and orient the cameras. The goal of positioning and orienting the cameras is to maximize the volume that can be tracked. For the calibration process (described in the next section) a central point in the volume has to be visible to all cameras. A minimum of two cameras have to see a marker in order for the software to be able to estimate the 3D location of that marker. Some experimentation is required to find a good configuration of the cameras. The configuration that worked best for us was to mount half the cameras at the top of the room and the other half at the bottom of the room. Cameras mounted on top face downwards to cover the floor area, whereas cameras that are mounted near the ground face upwards to cover the top area. Crossing the camera volumes in this way increases the number of cameras that can see a single marker. The crossing of camera view volumes and the resulting tracking volume is illustrated in **Error! Reference source not found.**.

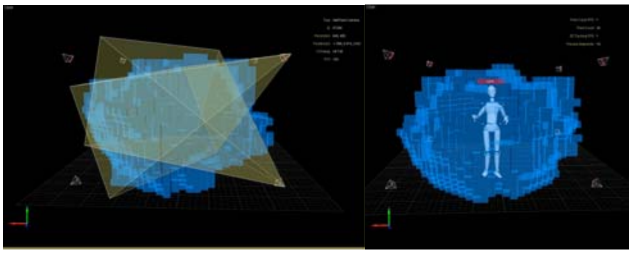

Figure 2 Left: Crossing camera view volumes to increase number of cameras that see a marker. Right: A skeletal representation of a participant in the viewing volume.

3.1.4 Arena Motion Capture Software

The Optitrack system includes a free SDK that enables a developer to access the cameras at a low level and to design his or her own motion capture software. In addition, a complete

[3] For example http://www.organicmotion.com/

[4] For example http://www.fourdviews.jexiste.fr/

[5] For example http://www.human-solutions.com/ and http://www.tc2.com/

system for tracking of individual markers and tracking of rigid bodies (tracking tools) is offered for less than $500. The Arena whole body motion capture software that currently can capture the movements of up to 2 human bodies with a 34 marker skeleton costs about $2000.

The motion capture system consists of 4 subsystems, a camera calibration system, a skeleton adaptation system, a motion recording and streaming system and finally a motion editing system.

Camera calibration is performed by moving a single marker in the tracking volume, in such a way that it is visible to multiple cameras while it is moved. The goal of the camera calibration is to find intrinsic and extrinsic properties of the camera. Intrinsic properties describe the distortions of the image caused by the camera's lens. Extrinsic properties specify the location and orientation of the camera in space. A trajectory consisting of a few hundred marker positions that are seen by a minimum of two cameras at a time is sufficient for the software to find a least squares solution in less than 5 minutes.

The Arena system supports two default skeletons, one can track 34 and the other 38 markers. The length of skeletal segments is adjusted to closely approximate those of the person to be tracked. This is achieved in a completely automatic way by the software which can be refined manually if required.

The Arena system can provide the processed skeleton motion data in real time over a network protocol to client software on the same or a remote machine. The streaming motion data can be provided in different forms: as joint positions, as absolute joint rotations or as hierarchical joint rotations and a root joint position. Rotation data is provided as quaternions.

The motion editing subsystem is specifically targeted at cleaning the motion data in an offline manner from artifacts that originate from incorrect identification or missing markers.

3.1.5 Markers and Body Suit

The marker based system uses spherical retro-reflective markers that can be identified by the cameras. The user wears a tight fitting body suit that is provided by the NaturalPoint company. Markers are placed on the body suit in a configuration that is defined by the software.

3.2 Head Tracking

The Optitrack system delivers stable tracking data for the whole body. We could therefore also use it for head tracking. However, we have experienced that during longer sessions the Optitrack system can occasionally introduce some jitter in the tracking probably owing to temporal occlusion of a marker. This is not a problem for the whole body, but proved to be distracting for the head tracking. We therefore are using the very robust and stable tracking data of an Ultrasound and Ineria based PC IS900 Intersense tracker. The head tracking data is streamed via VRPN[6] to our VR and character animation system.

3.3 HALCA Character Animation Library

For the animation and visualisation of the virtual body we use the hardware accelerated library for character animation (HALCA) [13]. HALCA uses the Cal3D XML file format [Cal3D06] to describe skeleton weighted meshes, animations, and materials. The core of HALCA consists of a motion mixer and an avatar visualisation engine.

The main goal of HALCA is to allow the user to visualise and animate several up to hundreds of very realistic looking virtual characters on single display PCs, in HMDs but also projection walls and CAVE like systems.

[6] http://www.cs.unc.edu/Research/vrpn/

3.3.1 Visualisation Engine

HALCA can be used in different rendering modes. In its simplest mode it uses basic OpenGL and so it runs on any graphics card that supports OpenGL potentially also on mobile devices.

HALCA can also exploit the latest GPU technology by allowing the hosting application to dynamically change GLSL shader programs [12]. This is illustrated in Figure 1 in which the avatar is visualised with dynamic shadows and mirror reflections and also with fairly good skin detail, using diffuse, bump, gloss and subsurface scattering maps to approximate real skin and clothing reflectance.

In addition to the usual shading in HALCA vertex shaders are used to perform the deformation of an avatar's skin according to the skeletal state of the avatar or by morph targets. Owing to the highly parallel nature of this problem current graphics hardware can carry out the required computations much more efficiently than the CPU. In addition, much less data is transferred between the CPU and the GPU. This is a key feature if large crowds of realisitic looking avatars are to be displayed. This is even more important if the display consists of multiple projectors that are driven by a networked render cluster [10].

During the initialisation, the mesh information along with morph target information is transferred to OpenGL Vertex Buffer Objects (VBO)s on the GPU. To save video memory HALCA reuses vertex and image map data if multiple avatars are visualised.

For animation in shader mode HALCA only transfers the skeletal joint transformations from the CPU to the GPU either as transformation matrices or as dual quaternions [7].

3.3.2 Animation Mixer

For avatar animation HALCA extends Cal3D's abstract mixer class by adding functionality to play, morph and blend animations, and in addition to directly access and modify the state of the whole or parts of the avatars skeleton efficiently.

Owing to the simple access to skeletal state of the character several Inverse Kinematics algorithms have been created for HALCA in the S3D scripting language and in C++ [11]. It is also possible to efficiently map tracked joint rotation data from a mocap system as described next.

3.4 Streaming Motion to HALCA

By using the NatNet libraries provided by NaturalPoint we developed a client dll that delivers the streamed data from Arena to HALCA. Since the avatar visualized in HALCA may have different skeletal dimensions as the one of the participant we carry out adaptation of the rotational joint information (based on an initial T-Pose of the participant) in order to match the motions of the Avatar with those of the participant.
The skeletal configuration is then passed directly to the GPU of the visualisation PC where the skin shader deforms the avatars skin and clothing to match the body posture of the participant. The avatar is then displayed in our wide field of view HMD.

3.5 Wide FOV HMD

The HMD we are using the wide field of view HMD Wide5 from FakeSpaceLabs[7]. This HMD needs to render the scene 4 times, both for left and right it requires a focused view and a peripheral view. These views are specified by the perspective matrices of the VR system.

[7] http://www.fakespacelabs.com

4. Use of the system

We are running several body ownership studies, which also confirm the robustness and functionality of our system. In these studies the participants carry out movements that are tracked by the system and mapped onto a life size avatar that visually replaces and/or mirrors the body of the participant (see Figure 1). Participants are free to move in any way and they are asked to observe their movements as carried out by the avatar. Our participants have commented on the very fast response time of the system and that even fast motions are tracked consistently. However, if a user partially leaves the tracking volume, then the motion capture software obviously cannot figure out the correct skeletal state of the user. In such situations the motion capture system may requires a few seconds to return to the correct skeletal state after the user user reenters the volume. This delay is caused by the passive marker system in which markers have no unique IDs. In our experiments this is not a problem because we ask our participants to stay within the volume. At the moment the adaptation of our skeletal information from Optitrack to HALCA is simple and works well. However, we are working on algorithms to more precisely match movements of real and virtual bodies.

5. CONCLUSION

We have described a system that enables us to virtually replace the whole moving body of our experiment participants in real time. We are using a low cost motion capture system that we integrated with our character animation and VR system. We have found good configurations of our cameras that allow us to track the whole body of our participant in a volume of about 20 cubic metres using 12 cameras.

Given that the user is within the tracking volume the system can deliver low latency full body avatar control. Because of the flexibility of the avatar animation and visualisation system of HALCA it is possible to dynamically change the appearance of the virtual character in terms of geometry and reflectance. For example it is possible to change the gender of a virtual body or to dynamically manipulate the reflectance of skin and clothing or the size of body parts. With this system it is possible to explore the transforming powers of VR and it is possible to experimentally explore in great detail brain functions that are responsible for the representation of our human body.

6. ACKNOWLEDGMENTS

This research has been funded by the EU FET PRESENCCIA project Contract Number 27731.

7. REFERENCES

[1] Matthew Botvinick and Jonathan Cohen. Rubber hands ‘feel’ touch that eyes see. *Nature*, 391(6669):756–756, 1998. 0028-0836 10.1038/35784 10.1038/35784.

[2] Marcello Carrozzino, Franco Tecchia, Sandro Bacinelli, Carlo Cappelletti, and Massimo Bergamasco. Lowering the development time of multimodal interactive application: the real-life experience of the xvr project. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 270–273, 2005.

[3] H Henrik Ehrsson. The experimental induction of out-of-body experiences. *Science*, 317, 24.8 2007.

[4] H Henrik Ehrsson, Charles Spence, and Richard E Passingham. That’s my hand! activity in premotor cortex reflects feeling of ownership of a limb. *Science*, 305:875–877, 2004.

[5] Radu P. Horaud, Matti Niskanen, Guillaume Dewaele, and Edmond Boyer. Human motion tracking by registering an articulated surface to 3-d points and normals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):158–164, January 2009.

[6] Simon Hay, Joe Newman, and Robert Harle. Optical tracking using commodity hardware. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, 2008.

[7] Ladislav Kavan, Steven Collins, Jiri Zara, and Carol O’Sullivan. Skinning with dual quaternions. In *2007 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 39–46. ACM Press, April/May 2007.

[8] Binga Leggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 8 2007.

[9] Betty J. Mohler, Heinrich H. Bülthoff, William B. Thompson, and Sarah H. Creem-Regehr. A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *APGV ’08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, page 194, New York, NY, USA, 2008. ACM.

[10] Guiseppe Marino, Franco Tecchia, and Massimo Bergamasco. Cluster-based rendering of complex virtual environments. *The 4th International INTUITION Conference on Virtual Reality and Virtual Environments*, 2007.

[11] Jesper Mortensen, Insu Yu, Pankaj Khanna, Franco Tecchia, Bernhard Spanlang, Guiseppe Marino, and Mel Slater. Real-time global illumination for vr applications. *IEEE Computer Graphics and Applications*, 28:56–64, 2008.

[12] Randi Rost. *OpenGL(R) Shading Language (2nd Edition) (OpenGL)*. Addison-Wesley Professional, 2006.

[13] Bernhard Spanlang. Halca hardware acclerated library for character animation. Technical report, event lab, universitat de barcelona, event-lab.org, 2009.

[14] Mel Slater, Bernhard Spanlang, Olaf Blanke, and Maria Victoria Sanchez-Vives. First person experience of body transfer in virtual reality. *Plos One*, (minor changes), 2010.

[15] Mel Slater and Martin Usoh. Body centred interaction in immersive virtual environments. In N Magnenat Thalmann and D Thalmann, editors, *Artificial Life and Virtual Reality*. John Wiley and Sons, 1994.

[16] Mel Slater and Martin Usoh. Representation systems, perceptual position and presence in virtual environments. *Presence-Teleoperators and Virtual Environments*, 2(3):221–234, 1994.

[17] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popovic. Árticulated mesh animation from multi-view silhouettes. *ACM Trans. Graph.*, 27(3):1–9, 2008.