Full Body Avatar Interaction

Bernhard Spanlang[1,3] and Mel Slater[1,2,3]
[1]Departament de LSI, Universitat Politècnica de Catalunya
[2]ICREA
[3]EVENT Lab, Universitat de Barcelona
www.event-lab.org

Abstract. This paper is a practice and experience paper on the use of a low cost motion capture system. We describe the integration of the Optitrack motion capture system with HALCA a hardware accelerated library for character animation in a Virtual Reality system called XVR. We demonstrate our system by mapping in real-time the tracked whole body motions of a person onto a virtual character which is visualized on an immersive large screen power-wall display.

Keywords: Motion capture, character animation, body perception, body scanning.

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. See, for example, [16, 17]. This has also come to be recognized more recently with a new wave of interest in HMDs, for example [10]. A virtual body representation has many applications not only for virtual reality and computer games but even in health care. However, in more common human-computer interaction, such as for computer games such (e.g., Quake) it is normal that no virtual body is required because the interactive images are shown from a first 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 comes from 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

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

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 virtual body part or even a whole virtual body [9, 4] belongs to the human participant. The rubber hand experiment was an early starting point of research in this direction [1]. Ehrsson et al [3, 5] showed that particular brain areas are activated if a human participant has “the rubber hand experience”.

This paper describes a system that we built in order to carry out further experiments on the impact of a virtual body on the sense of body ownership. The system consists of a low cost optical tracking system, a high resolution immersive stereo display system, software for whole body tracking, a graphics hardware accelerated library for character animation (HALCA) [15] and the XVR Virtual Reality 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.

2 Related 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 if visualization of a deformable body is desired. 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 [18] 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 [6] 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.

---

[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/

Pintaric and Kaufmann [13] present a low cost marker based tracking system that uses multiple Firewire Cameras. 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 [7]. With current restrictions of the Wii mote this system allows to track only 4 markers but with less than 100€ the system costs one to two orders of magnitude less than any other 6DOF tracking system. The systems described in the papers cited above use a camera calibration technique similar to the one described by Zhang et al in [19]. 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 two main components, a low cost full body motion capture system and a hardware accelerated library for character animation (HALCA). In the following subsections we detail these systems.

3.1 The Optitrack full body motion capture system

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, the cameras and software used.

3.1.2 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 contain much infrared and therefore the system works well with such artificial light. However, the system is very sensitive to specular 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 bounced by a specular surface are difficult to mask and therefore have to be avoided as much as possible in the scene.

3.1.3 Cameras

Optitrack V100 cameras consist of a 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.

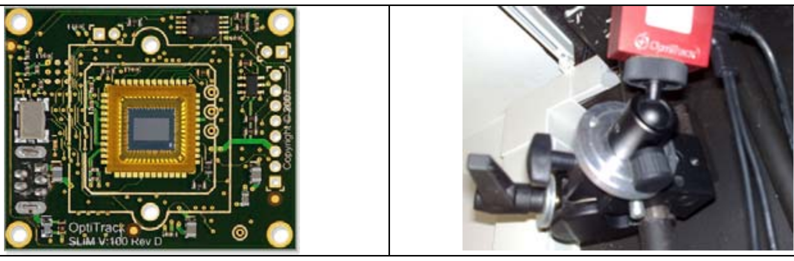

Figure 1 Left: Optitrack CMOS chip on PCB board, Right: Flexible Manfrotto camera mount.

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 some threshold method on the camera so that only 2D marker positions have to be 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 camera-mounts that enable us to freely position and orient the cameras (see Right of Figure 1). 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 by 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 was 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 is illustrated in Figure 2.

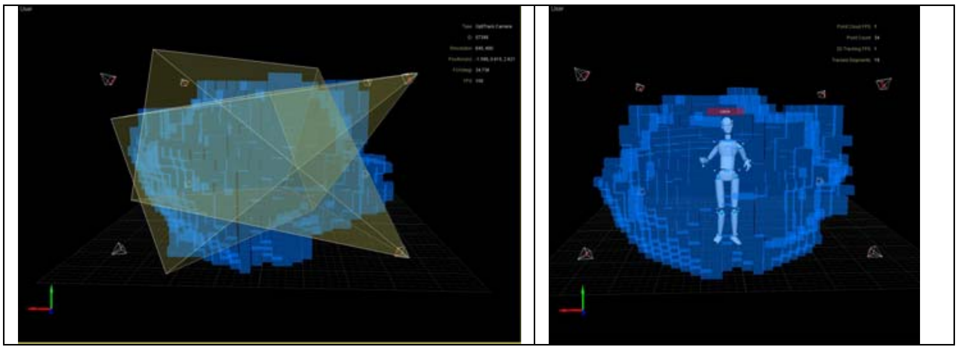

Figure 2 Left: Crossing viewing volumes to increase number of cameras that see a marker. Right: A skeletal representation of a character in the viewing volume.

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 system for tracking of individual markers and tracking of rigid bodies (tracking tools) can be purchased 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, the camera calibration system, the skeleton adaptation system, the motion recording and streaming system and finally a motion editing system.

Camera Calibration System

Camera calibration is achieved by moving a single marker in the tracking volume, in such a way that it is seen by most of the cameras while it is moved. This process is called “Wanding” in Arena and in the Tracking Tools software. The goal of the camera calibration is to find intrinsic and extrinsic properties of the camera [19]. 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 work out the required properties in less than 5 minutes. The solution of this process is achieved by a proprietary optimization process that minimizes a least squares error.

Skeleton adaptation

The Arena system supports two default skeletons, one can track 34 markers and the other can track 39 markers. The length of the skeletal segments has to be 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.

Motion Streaming/Recording

The Arena software can provide the processed skeleton motion data in real time over a network protocol to a client 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. Plugins for 3D Studio Max Character Studio and Motion Builder are provided by Natural Point to enable the user to directly stream motion data into these modeling softwares.

Motion Editing

The motion editing subsystem is specifically targeted at cleaning the motion data from artifacts that originate from incorrect identification of markers once a complete motion sequence was recorded. Specific tools are available that enable the user to mark areas of such recorded motion data in which the user can for example swap marker information to fix problems where a marker id was confused with another or to let the software automatically fill temporal holes in the data that are caused by markers that could not be located for small periods during the capture.

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 HALCA Character animation library

For the animation and visualisation of avatars we are using the hardware accelerated library for character animation (HALCA) [15]. 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 animate and visualise several up to hundreds of realistic looking characters on single display PCs, in HMD (head mounted displays) and CAVE like systems. HALCA is closely related to Cal3D, extends Cal3D's animation and visualisation functionalities and allows the user to script it from VR applications.

3.2.1 Visualisation Engine

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

When HALCA is run in shader mode it can either load and activate GLSL shader programs [14] from a text file or it assumes that a shader program of the OpenGL context was activated by the hosting application.

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 very important feature when it comes to visualising larger crowds of realisitic looking avatars. This is even more important if the display system consists of multiple projectors that are driven by a networked render cluster [11].

During the initialisation, when avatars are loaded the mesh information along with morph target information is loaded into OpenGL Vertex Buffer Objects (VBO)s on the GPU. HALCA has functionality to reuse vertex and image map data. For example if two avatars share the same mesh but have different textures then the same VBO is used for both characters with the corresponding texture maps. This functionality is very useful if a bigger number of similar avatars is required.

For animation in shader mode HALCA only has to transfers the skeletal joint transformations from the CPU to the GPU either as transformation matrices or as dual quaternions [8].

3.1.2 Animation Mixer

For avatar animation HALCA extends Cal3D's abstract mixer class and adds the following functionalities:

- play and blend animations of different avatars at different speeds
- play and blend temporal parts of an animation.
- play or morph an animation only on specified body parts.
- go through an animation not by time but by an external value. We call such animations morph animations not to be confused with morph targets.
- directly access and manipulate joint rotations and translations on top of blend and morph animations.
- efficiently access and manipulate the whole skeletal state (One function call is required to accessed the skeleton or to set it to a particular pose by passing the skeletal state vector of an avatar to or from HALCA). This is useful for example for the integration with a real time whole body motion capture system. Efficient access to the skeletal state is also useful for example for a physics engine to check and respond to collisions between avatars or limbs of an avatar.

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++ [12].

3.3 Streaming between Optitrack and HALCA

By using the NatNet libraries provided by NaturalPoint we developed a client dll that can deliver the streamed data from Arena directly to HALCA. A complete system overview is shown in Figure 3. Since the avatar visualized in HALCA may have a different skeleton to the skeleton in the Optitrack system we carry out adaptation of the rotational joint information in order to match the skeletal motions of the Avatar.

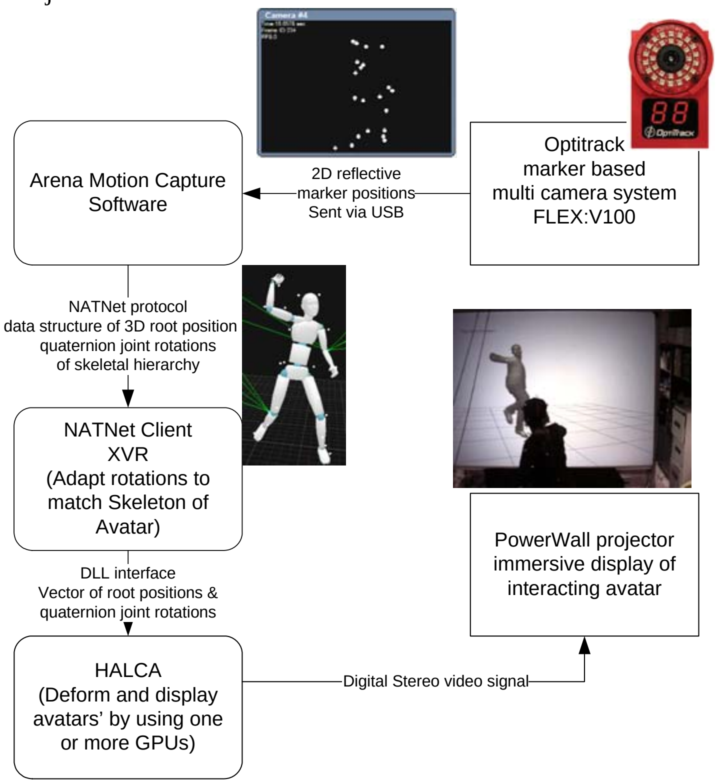

Figure 3 System data flow.

Once HALCA knows the skeletal configuration this information is passed on directly to the GPU where the appropriate shader accordingly deforms the avatars skin and clothing. In the current configuration we visualize the avatar on a large screen stereo power wall.

4 Results

We have run several pilot studies the goal of which was to identify the robustness and functionality of the system. In these pilots the participants would carry out movements that were tracked by the system and copied by an avatar that was displayed as a life size human on an immersive power-wall in front of the participant. The participants were free in choosing their movements and they were asked to observe the related movements that the avatar carried out. Our subjects have reported, that the system has a very fast response time and that even fast motions are tracked consistently. However, if a user partially leaves the tracking volume, then the Motion

Capture software obviously has difficulties in figuring out the correct skeletal state of the subject. In such situations even if the subject re-enters the tracking volume completely the motion capture system may require some time to adapt to the correct state. This delay is caused by the fact that the Optitrack system is a passive marker system and therefore markers do not have unique IDs. The system finds the markers based on the known skeletal configuration.

At the moment the adaptation of our skeletal information from Optitrack to HALCA is simple and does not take into account the influences of different length of skeletal segments. Therefore if for example the user claps his hands then the avatar displayed may not join the hands if the Avatars skeleton is very different to that of the user.

5 Conclusions and Future Work

We have described a system that enables the user to control a humanoid avatar in real time by moving the whole body. We are using a low cost motion capture system that we integrated with our character animation and virtual reality system. We have found good configurations of our cameras that allow us to track the markers well.

Given that the user is within the tracking volume the system can deliver low latency avatar control that mirror the motions of the user.

Next we are going to improve the behaviour of the system to quicker recover from situations when some markers are not within the tracking volume. In addition we would like to improve the adaptation of the skeletal rotations between tracked skeleton and the skeleton of the avatars, so that the avatar can match the motions of the user better.

We are also working into extending the system so that the avatar does not mirror the users motions but acts in a way similar to that of two real people interacting in a conversation.

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

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, Nicholas P. Holmes, and Richard E. Passingham. Touching a rubber hand: Feeling of body ownership is associated with activity in multisensory brain areas. *Journal of Neuroscience*, 25(45):10564–10573, 2005.

[4] H Henrik Ehrsson. The experimental induction of out-of-body experiences. *Science*, 317, 24.8 2007.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] Binga Leggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 8 2007.
[10] 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.
[11] 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.
[12] 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.
[13] Thomas Pintaric and Hannes Kaufmann. Affordable infrared-optical pose-tracking for virtual and augmented reality. In *Trends and Issues in Tracking for Virtual Environments*, Charlotte, NC, USA, 2007.
[14] Randi Rost. *OpenGL(R) Shading Language (2nd Edition) (OpenGL)*. Addison-Wesley Professional, 2006.
[15] Bernhard Spanlang. Halca a library for presence research. Technical report, event lab, universitat de barcelona, event-lab.org, 2009.
[16] 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.
[17] Mel Slater and Martin Usoh. Representation systems, perceptual position and presence in virtual environments. *Presence-Teleoperators and Virtual Environments*, 2(3):221–234, 1994.
[18] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popovic. Ärticulated mesh animation from multi-view silhouettes. *ACM Trans. Graph.*, 27(3):1–9, 2008.
[19] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.