Real-Time Motion Tracking in digital art installations

Oliver Habibi

MSc (Design & Digital Media)

Table of Contents

Introduction

Over the past fifteen years, motion capture (the interpretation by computer of recorded movements) and motion tracking (live interpretation of movements by computer) have become increasingly accepted and accessible as a method of accurately plotting organic movement into computers to bypass the tricky and cumbersome process of manually animating a 3D model with realistic movements.

The list of potential uses for motion capture does not stop there, and continues to rise as increasing computational power become available, alongside a general acceptance of motion capture into the mainstream. Linking it with other technologies, such as real-time graphics manipulation, and using motion tracking in novel ways, such as a sole interface to an art installation, can have interesting results. As the technology and the systems it relies on mature and become widely accessible, its usefulness is increasing exponentially.

The current acceptable quality of motion capture and motion tracking mark a clear milestone in the long quest to create an intelligent machine; a computer can now 'see'. These advances have been largely in tandem with progress in the digital equivalents of feeling and hearing, and are providing benefits in a wide array of industries. Sports scientists can study the biomechanics of the body in motion in more detail than ever before, and animators can bring their models to life with realistic motions that maintain the suspension of disbelief for film watchers or computer game players.

While these more technically savvy (and importantly, well-funded) sectors of the entertainment industry have embraced motion capture and tracking with open arms, there are other elements who have overlooked or even avoided this technology: performance art and installation art are two such examples. The few trailblazers who have tested the water to date have often come up with remarkable results – a human dancer interacting with

a digital one[1], and an 'embodied conversational agent' (computer-modelled head projected onto a screen, speaking in reply to typed audience questions) with "real-time lip-synching, speech synthesis and facial expressions"[2], to name but two.

The creators of the above examples (the Genoa Opera House and Stelarc), or at least their collaborators, are notable for being well-funded technophiles; not a common attribute among all the artists generally associated with dramatic, dance and installation work. They are often willing to harness current technology, but would perhaps view motion-capture as esoteric and beyond their budgets, and with good reason. However, this essay will show that affordable and competent motion tracking methods are emerging, enabling whole new approaches in interactivity that are waiting to be explored.

Coming at this mix of theatre and technology from another direction, the MSc (Design and Digital Media) and BMus (Music Technology) students of 2002-2003 will be holding an installation (an exhibit of artistic work in the surroundings of a gallery) of manipulated photography during the week of Monday 12th May 2003 at the Matthew Gallery in Edinburgh. Audience interactivity will be a feature of the installation – viewer movements will be traced around the installation by a video camera using basic 2D motion tracking, giving two sets of input via midi technology to the software controlling the installation. The values and variations of these two inputs will control the display of images, which will respond to viewer movements.

Even for this rather basic motion-tracking environment, the expertise of two different postgraduate courses has been required, with budget not a major consideration as software licences and the installation space are being provided free of charge, and none of the twenty contributors are being paid a fee for their work.

So is motion capture, and specifically motion tracking, just a gimmick and a minefield for digital or semi-digital installations and performances? What

are it’s limitations and capabilities, how will it progress, and how vital will it become in the future?

Motion Tracking and Digital Installations

Motion Tracking vs. Motion Capture

Motion tracking is the real-time equivalent of motion capture. Just as the real-time and responsive graphics of contemporary computer games compare to the photo-realistic, pre-rendered and non-interactive images that are prepared over months and integrated into modern Hollywood films, motion tracking is relatively constrained by the computational power available to it. Advanced are constantly being made; with an actor wearing a special suit and sporting seventy-one strategically-placed sensors (thirty on the face, forty-one on joints and body extremities), a useful medium-polygon model can currently be animated in real-time with audio on one powerful workstation[3].

The exact power of the workstation being used is important, as it will always bottleneck the grandeur of available real-time effects. Advances in processing power, increases in memory size and bandwidth speed, and hardware video implementation of graphics libraries like DirectX or OpenGL are all extremely welcome. Still, utterly convincing photo-realistic scenes that are fully editable and interactive in real-time are a long way off. This is well illustrated by the figures from WETA, the digital effects studio responsible for computer graphics scenes in the current 'Lord of the Rings' set of films. For rendering the unparalleled graphical sequences, 192 Dual Pentium 1 GHz and 448 Dual 2.2 GHz processors, a total of 1280 processors running at approximately 2,355 GHz[4] worked non-stop, twenty-four hours a day for months, producing the one-thousand shots in the "Two Towers" episode of the trilogy that featured CGI. Real-time has never looked so distant!

The History of Motion Tracking

The need to capture motion has been identified for decades in various fields, many of them unsurprisingly with no direct relation to character animation, interactivity or digital media - for instance, intruder detection systems. However, any advances in these industries go unaccredited in documented histories of motion capture, which prefer to point to a method called 'rotoscoping', first used by Walt Disney, who traced animation over film footage of live actors playing out the scenes of the cartoon 'Snow White'[5]. The quality of animation produced is high, but as the tracing must be done by hand, many of the advantages of automated motion capture are lost.

The first real motion capture took place in the early 1980s, with potentiometers being attached to a moving body, and the generated vectors and coordinates being used to animate computer figures to clinically study movement abnormalities[6].

Only after that did the computer graphics industry become involved, and attempt to optically track human movements using small markers (flashing LEDs or reflective dots) and multiple cameras to plot the 3D coordinates of the markers[7].

In the late 1980s, the idea of a prosthetic body suit or exoskeleton full of sensors that actors could wear while mimicking the needed movement was explored[8].

These techniques have consistently been refined, the exponential increase in computing power over time enabling more data to be scanned per second, giving more realistic movement all the time. As was stated previously, processing power is very important; at this moment in time, details down to facial muscle movements can be captured in real-time, but there is still much room for improvement!

Different Types of Motion Tracking

There are a few different methodologies when it comes to capturing data in real-time. The three main techniques are prosthetic, magnetic and optical.

Prosthetic Motion Capture

Prosthetic motion capture uses potentiometers on the aforementioned plastic exoskeleton that an actor must 'wear', and then act out his or her movements. This technique is obviously only of use in humanoid character animation, but is very accurate and transmits real-time data at a far greater range than any other technology. The suit is cumbersome but its advantages mean that prosthetic motion capture has thrived.[9]

Figure 1: the Gypsy 3 prosthetic full body suit

Magnetic Motion Capture

Using magnetic motion capture, sensors attached to the body being animated are manipulated inside a magnetic field. This technique is the least power-hungry in terms of computational number crunching, so is the closest to real-time, with up to one hundred samples a second possible. The sensors also provide details on their orientation in full 3D.

Figure 2: a suit with magnetic sensors in a closed environment

One of the few caveats is the obvious effect that any metal would have on the generated magnetic field, the only sure way to avoid this would be to use a dedicated motion-capture studio. Magnetic motion capturing also

requires a very tight space to be used, with a range of only three metres[10].

Optical Motion Capture

Optical motion capture systems do away with a suit or exoskeleton, and do not necessarily even need physical markers such as the LED lights or reflective dots – systems exist that are capable of mapping patches of colour or brightness to areas on a 3D model in real time (for instance matching the eyes of the actor to the eyes of a 3D gladiator model etc).[11] Optical trackers also generally have a higher range than magnetic ones, as they can use any camera video feed. Perhaps the greatest advantage of optical motion capture is that it is not limited to the movements in a closed motion-capture studio; any movement on a video stream, live or recorded, can be analysed… water flow in a river, traffic speeds on motorways, the list is endless.

Figure 3: Optical facial tracking (2 cameras)

Unfortunately there are also several big disadvantages. In a character animation system, some markers will not be visible to the camera, or set of cameras, at any one point, and their position must effectively be guessed. Trackers may also be too close together for the system to register a difference. And while triangulation using multiple cameras can build up a 3D model from the assorted 2D video streams, the orientation of the sensors is not picked up. Computational power must make up for these shortfalls, meaning that real-time accuracy must be sacrificed in order to keep up with the actor movements. For precise results, post-processing is invariably needed to adjust any wayward signals.[12]

Expanding on a point in the last paragraph, video streams deliver 2D images a number of times per second – there is no sure way of accurately calculating movement along the Z-axis without using multiple camera that

are exactly calibrated to enable 3D plotting of scene elements. With no easy means of calculating object orientation either, the set of data from a simple optical motion capture system will be lacking compared to magnetic or prosthetic techniques. Only X and Y coordinates for objects that are in unobstructed view of the camera can be derived in real time.

For the purpose of live performance or audience interaction such as in the Edinburgh installation, wearing a cumbersome prosthetic or sensor-laden suit are out of the question – and optical motion tracking is required by default. The lack of 3D data on visitor movement is inconsequential as only two user inputs are needed; left-right and forwards-back in relation to the installation.

Cost - The Low End

Set-ups such as the particular multi-camera, high frame-rate, 1.3-megapixel example at the beginning of this section are at the high end of both available technology and cost. At the other end of the spectrum, motion-tracking software can be surprisingly affordable, with SmoothWare Design's 'TrackThemColours' Xtra for Macromedia Director downloadable for $150, giving the popular multimedia application the ability to track objects in a video clip by colour, brightness or pattern, and return 2D or 3D coordinate values for real-time calculations[13]. Other examples are the free 'WebCamXtra', supporting motion detection, colour tracking, glob distinction and pixel addressing at 160x120 resolution[14]; 'Cyclops', an $99 object for the Max/MSP program that analyses QuickTime video and tracks colour movement[15]; BigEye is a $150 motion-tracker that outputs midi signals[16]. A final and fitting example is SoftVNS, a powerful $350 Max/MSP object that can analyse video in real time, which the author uses in his own installations as an interface to a digital environment[17] (See figure 1).

Figure 4: SoftVNS in action - all motion from the video stream on the right is filtered out and shown on the left

Together with a licence for Director or Max/MSP, a digital video camera and a modern PC or Macintosh computer, a bona fide motion tracking system can be had for as little as £2000 – the price of a middle range projector and screen, so ubiquitous in installations. The software and hardware used for the Matthew Gallery Installation will be of this order, giving simple 2D plotting of peoples' movement around the space, with no gesture recognition or particular interpretation of the data.

Cost - The High End

The figures and complexity of the high-end, dedicated motion capture and tracking facilities are far higher; Edinburgh University's own Virtual Environment Centre (EDVEC) received initial funding to the tune of £610,000[18] and provides motion capture including gait analysis (the biomechanical study of how people walk) and 3D scanning[19].

Figure 5: Edinburgh University Virtual Environment Centre

With construction costs of this magnitude, rental rates to hire out a motion capture laboratory are never going to be affordable to many smaller outfits interested in using the technology, never mind the additional costs of hiring actors and other essentials.

Cost - The Mid Range

In between the inexpensive Max/MSP objects and the massive investment needed to build a motion-capture studio, lie the vast majority of motion tracking solutions. From Maya plugins like 'Eyematic FaceStation' at $2000, an optical face-scanning program with no sensors needed[20] – to the similarly priced 'DIEM Digital Dance System' software/hardware mix dedicated to interpreting dance movements[21], the choices on offer are slightly more complex and esoteric than the cheaper options, and give more accurate results.

Motion Tracking as an Interface

One interesting use of the data provided by motion tracking systems (2D or 3D coordinates, orientation angles, etc) is as set of inputs to an interactive computer program. Specific examples and methods will be looked at later, but for now it is worth noting that the software, hardware and human audacity needed have all come together to provide the means of creating and experimenting with real-time motion tracking as an interface.

There are some important considerations though; the process must be simple enough for an audience to appreciate the connection between movement or gesture and the system's output, but not too simple for too long, as the novelty will quickly wear off[22]. After all, this is just a new trick made possible by technology, not an artistic statement all by itself.

Improvisation

An interesting side effect of the introduction of audience participation into an installation is the necessity of improvisation in the final product[23]. The classical audience behaviour and input models for installation performances can be turned on their head. How, though, can an audience-influenced piece be guaranteed as compelling to the same audience? Over-reliance on motion tracking as an interface without due consideration to the material being manipulated by the interfacing will surely be detrimental to the whole, and negate the forward strides taken with technology by ignoring the fundamentals of art.

Motion capture in the Matthew Gallery Installation and Other Examples

The Matthew Gallery Installation

Motion capture is often a daunting concept, but the imminent exhibition in the Matthew Gallery, Edinburgh, will use a relatively simple implementation of motion tracking to construct an intriguing method of interfacing with the interactive content of the installation.

Viewers will enter a room and see a set-up of three large perpendicular white screens displaying back-projected images, in the shape of 3 walls of a square room. The three displays will be touching, giving an impression of three connected images, a triptych. A camera will be positioned, enabling it to pick up movement around the space.

The triptych system has been developed using a mouse to interface with the software controlling the images. Mice provide vertical and horizontal positioning at a rate between 20Hz (serial mice) to 120Hz (USB mice) and the human eye's persistence of vision threshold is around fifteen frames per second. The camera will have to provide a similar rate of information to do justice to the smooth transition effects that the installation uses.

The camera feed is analysed in the Max/MSP software on a Macintosh G4 system, and the perceived viewer coordinates are converted to midi signals and sent to another G4 for interpretation by Macromedia Director, using an appropriate Xtra (plugin) to decipher the signals into two variables. The camera has replaced the mouse completely as an interfacing tool! Director will then control the triptych and ambient sound to reflect the movements around the space by the viewer.

It all sounds very complicated, but has been developed from scratch in five weeks by the installation artists.

Sensuous Geographies

A comparable but grander example of the same concept is 'Sensuous Geographies' by Alistair MacDonald and Sarah Rubridge. For a full description of the piece, please see appendix 1. Motion tracking is used to 'attach' sounds to protagonists (including audience members) who move about a space, wearing a brightly coloured gown and hat, controlling their sound through various parameters such as volume, pitch and stereo placement[24]. The result is an incredibly engrossing and atmospheric sonic experience, with the technology safely encapsulated away from an empowered audience; years of research and tens of thousands of pounds of funding delivering a powerful final product.

Figure 6: Sensuous Geographies

Conclusion

Motion capture is an intriguing emerging technology, the benefits of which are beginning to trickle down from the high-spending film industry to the small market of digital art installation.

Of all the motion tracking techniques, optical tracking seems the most likely candidate for use in live performance, as an unwieldy plastic suit need not be worn if the camera and recognition software are calibrated well. Optical motion capture is advancing in speed and accuracy as computers in general become more powerful, and while the horsepower for intricate real-time video analysis and 3D coordinate construction from 2D images is not yet available, meaningful artistic works can now be undertaken with contemporary technology.

The technology may not even be prohibitively financially costly, but any artistic director or choreographer will greatly appreciate a computer programmer or technician in building and setting up a motion tracking system. Those in the fortunate position of having the necessary expertise are already uncovering an exciting new world of organic interfacing and live machinery.

Appendix 1

*"**Sensuous Geographies** is a responsive sound and video environment. It is a space of ritual, a liminal space intended to draw attention both to the intricate inter-weaving pathways of sensation in the body, and to the equally intricate pathways of subtle group behaviour.*

The installation comprises a collection of polyphonic electroacoustic sound worlds which are created as the actions of visitors to the installation generate and influence both individual musical layers and the overall texture of the sound environment. Shadowy virtual counterparts of the visitors are projected onto banners, intermittently brought to presence by the motion of those engaging with in the installation. The members of the public who become active participants in this responsive environment immediately become an integral part of the installation itself, as they don richly coloured costumes before entering the active space in order that they can be identified as distinct individuals by the motion tracking which drives the interactive system. The visitors' engagement with and immersion in the sound environment results in a continually evolving world of sound, image, light and colour which visitors can alternately observe from without and experience from within.

Sensuous Geographies is intended primarily to be a space for interaction, for contemplation, even for play. It is in a very real sense a performative space, for the users both generate and become an integral part of the environment as they engage with it. When activated the piece exists simultaneously as a durational 'performance' and an immersive installation environment[25]."

An excellent video showcasing the installation in use is available at:

http://mcdiddle.webspace.fish.co.uk/movie1.html

References

[1] Arnspang, Jens, "MOSART Task 4 survey", page 4. Available online at: *http://www.diku.dk/forskning/musinf/mosart/midterm/T4/arnspang-detection.pdf*

[2] Stelarc, 2003; "Prosthetic Head". Available online at: *http://www.newmoves.co.uk/archives/newterritories2003/stelarc.htm*

[3] Vicon Motion Systems, 2002; "Full body". Available online at: *http://www.vicon.com/entertainment/stories/nazlan.shtml*

[4] The One Ring website, 2003. Visited 5/5/03. Available online at: *http://www.theonering.net/perl/newsview/8/1047582857*

[5] Sturman, David, 1999; "A brief history of Motion Capture for Computer Character Animation", page 1. Available online at: *http://www.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/history1.htm*

[6] Sturman, David, 1999; "A brief history of Motion Capture for Computer Character Animation", page 1. Available online at: *http://www.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/history1.htm*

[7] Sturman, David, 1999; "A brief history of Motion Capture for Computer Character Animation", page 3. Available online at: *http://www.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/history1.htm*

[8] Sturman, David, 1999; "A brief history of Motion Capture for Computer Character Animation", page 3. Available online at: *http://www.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/history1.htm*

[9] Mellor, David, 2002; "A Sound Person's Guide to Video", Audio Media Magazine. Available online at: *http://www.spgv.com/columns/motioncapture.html*

[10] Mellor, David, 2002; "A Sound Person's Guide to Video", Audio Media Magazine. Available online at: *http://www.spgv.com/columns/motioncapture.html*

[11] Mellor, David, 2002; "A Sound Person's Guide to Video", Audio Media Magazine. Available online at: *http://www.spgv.com/columns/motioncapture.html*

[12] Sturman, David, 1999; "A brief history of Motion Capture for Computer Character Animation", page 2. Available online at: *http://www.siggraph.org/education/materials/HyperGraph/animation/character_animation/motion_capture/history1.htm*

[13] Rozin, Daniel, 2001; "TrackThemColors Introduction". Available online at: *http://www.smoothware.com/track.html*

[14] Nimoy, Josh, 2003; "Introductory WebCamXtra Tutorial". Available online at: *http://webcamxtra.jtnimoy.com/*

[15] Singer, Eric, 2003; "Cyclops by Eric Singer". Available online at: *http://www.cycling74.com/products/cyclops.html*

[16] Baldé, Frank; "BigEye". Available online at: *http://www.steim.org/steim/bigeye.html*

[17] Rokeby, David, 2002; "SoftVNS Highlights". Available online at: *http://www.interlog.com/%7Edrokeby/softVNS.html*

[18] Middleton, Roy, 1998; "EDVEC 3D Capture Studio" in "Graphics and Visualisation" issue 60. Available online at: *http://www.agocg.ac.uk/gv/issue60/edvec.htm*

[19] EDVEC Web page, visited 5/5/03. Available online at: *http://www.edvec.ed.ac.uk/html/services/index.html*

[20] EyeMatic, 2003; "EyeMatic Facestation", visited 5/5/03. Available online at: *http://www.eyematic.com/products_facestation.html*

[21] Siegal, Wayne, 2002; DIEM Digital Dance System web page. Visited 5/5/03. Available online at: *http://www.musik-kons.dk/diem/prices.html*

[22] Naugle, Lisa Marie; "Composition/Interactivity Issues" web page. Visited on 5/5/03. Available online at: *http://music.arts.uci.edu/dobrian/motiontracking/aesthetics.htm*

[23] Naugle, Lisa Marie; "Composition/Interactivity Issues" web page. Visited on 5/5/03. Available online at: *http://music.arts.uci.edu/dobrian/motiontracking/aesthetics.htm*

[24] MacDonald, Alistair, 2003; "Sensuous Geographies: a performative sound and video installation". Available online at: *http://www.sensuousgeographies.co.uk*

[25] MacDonald, Alistair, 2003; "Sensuous Geographies: a performative sound and video installation". Available online at: *http://www.sensuousgeographies.co.uk*