

# Part I

## Human Motion Capture



## Chapter 2

# Interacting with a Virtual World through Motion Capture

### Synopsis

In this chapter a general introduction to MoCap is given together with a general discussion of how MoCap can be applied in HCI. The chapter consists of a publication from 2001 [A] which has been published as chapter 11 in the book titled *Interaction in Virtual Inhabited 3D Worlds* [B]. The context of the publication is therefore how to apply MoCap to interact with a virtual world. Note that when there is a reference to *this book*, e.g. "Previously in this book...", *this book* refers to [B] rather than to this thesis.

### Synopsis References

- A. T.B. Moeslund. Interacting with a Virtual World through Motion Capture. In Lars Qvortrup, editor, *Interaction in Virtual Inhabited 3D Worlds*, chapter 11. Springer-Verlag, 2000
- B. Lars Qvortrup, editor. *Interaction in Virtual Inhabited 3D Worlds*. Springer-Verlag, 2000



---

## Interacting with a Virtual World Through Motion Capture

Thomas B. Moeslund

### 2.1 Introduction

Previously in this book the concepts of interactions, agents, avatars and virtual worlds have been discussed. In this paper we will look closer at how the non-verbal interaction with a virtual world takes place.

The non-verbal interaction (hereafter interaction) to virtual worlds can be through a command/text window and keystrokes which are seen in many systems. But as the virtual worlds and the interaction with them becomes more complex so does the need for more advanced interfaces. One way of improving the interfaces is to use motion capture (MoCap). That is, the motion of the subject interfacing to the virtual world is used to make the interface more general and intuitive. As an example, say you want to make a creature in the virtual world wave its hand. Instead of using playback, with its inherent limitations, or using a complex combination of keystrokes and mouse movements, you could simply wave your own hand in the desired way and have a MoCap device register the movements and send them to the creature, making it able to duplicate your movements in real-time. The advantages of using MoCap are evident, but how exactly can the motion of a human or another object be captured? And how may it be used to interact with the virtual world? These are the questions we will try to answer in this paper.

We will do this by first describing the different devices used for MoCap and their complexity, and then discuss how MoCap is used to control something, especially in a virtual world. But first we will give a more precise definition of MoCap, including a brief glimpse into its development.

### 2.2 Motion Capture

A fair question to ask in this context is what exactly is MoCap? Formally it is the process of capturing motion, usually human body motion, via a MoCap device. However, a closer investigation reveals that this is not entirely true. Many devices, e.g. a computer mouse or a steering wheel in a car, are usually *not* considered MoCap devices, and their tasks are not considered to be MoCap, even though their sole purpose is to capture human motion. A precise definition covering all aspects of MoCap cannot be given, since it differs between applications and depends on the

applied technology. Generally it can, however, be said that measurement devices which are associated with registering and/or storing of general human motion, and which do not have a predefined purpose (as, for example, a steering wheel does) are considered MoCap devices. Another characteristic is that most of these devices can be fitted to different humans and applications.

When we use the term MoCap in this chapter it refers to *the process of capturing the motion of a human body (part), at some resolution*. The phrase *at some resolution* is added to indicate that both the estimation of the overall human motion, e.g. his or her centre of mass, as well as the estimation of motion of each articulated joint in an object, is considered a MoCap problem.

The development in MoCap technologies and devices comes from two sources: research and industry. At first, for many years, MoCap applications were designed to measure, and thereby help understand, the motion patterns of humans (and animals). This was solely driven by research. Later medical aspects were investigated, resulting in an actual MoCap market.

Some 10-20 years ago, as computers were becoming widely available, a number of new application areas opened up and the interest in MoCap exploded. Since then some of the interested parties have been big industries: military, entertainment, medical and advertising. They all have plenty of money to invest, resulting in a fast developing market with a continuous flow of new and improved products. Beside these industries the research community has also contributed to the rapid development. Lately the research community has invested much effort in the area of Human Computer Interaction (HCI). For complex HCI applications, such as sign language recognition and avatar control in a virtual world, MoCap is a very welcome (and in some cases necessary) tool.

Together the research community and various industries are constantly developing and improving MoCap devices. The most important devices and their underlying technologies are described in the following section.

## 2.3 Devices Used for Capturing Motion

This section gives an overview of the different types of devices used for MoCap and a glimpse into the history of visual capturing methods in particular. The different MoCap devices are based on either active or passive sensing.

The concept of active sensing is to place devices on the subject which transmit or receive real or artificially generated signals. When the device works as a transmitter it generates a signal which can be measured by another device located somewhere in the surroundings. When it works as a receiver it receives signals usually generated by some artificial source in the surroundings.

In passive sensing the devices do not affect the surroundings. They merely observe

what is already in the world, e.g. visual light or other electromagnetic wavelengths, and generally do not need the generation of new signals or wearable hardware.

### 2.3.1 Active Sensing

Different devices for active sensing have over the years been invented to capture human motion in particular. Marey was the first to tackle the MoCap problem, as early as 1873. He used pneumatic sensors and pressure chambers, e.g. under the foot, to measure the pressure conducted by the foot [17]. This kind of work, where the forces of movements of different body parts are measured, is known as kinetic-based MoCap.

In the last two decades different devices, such as mechanical sensors, accelerometers, electromagnetic sensors, acoustic sensors and optic fibres, have been invented; see [3, 11, 13, 15, 28] for general descriptions of these sensors and [1, 2, 5, 21, 22, 26] for examples of systems where they are used in practice. In the following a short description of the different devices is given.

**Mechanic devices** are attached to some movable parts which when moved, e.g. during bending, will output a signal directly reflecting the configuration of the movable parts to which it has been attached. In figure 2.1 a glove (the CyberGlove) based on this technology is shown.



Figure 2.1: A mechanical glove for MoCap [3].

An **accelerometer** is a small device which measures the acceleration of the object it is attached to. This is done by measuring deflection caused by the movements of the device and converting this into an electrical signal. Devices can be more or less sensitive to the Earth's gravitational field.

**Electromagnetic devices** are attached to different joints/segments on a subject and measures the orientation and position of each device with respect to the Earth's magnetic field, or a field generated by a transmitter. In figure 2.2 a subject wearing

seven electromagnetic devices is shown. All devices feed their signals to a box on the back of the subject from where they are send (via radio or through a wire) to a MoCap computer.



Figure 2.2: Electromagnetic devices mounted on a subject to capture his movements.

**Acoustic devices** use a set of microphones to receive a sound wave emitted from an acoustic sensor attached to a subject, or visa versa. Using either triangulation or the phase of the sound wave it is possible to calculate the 3D position of the device.

An **optic fibre device** is a special case of the mechanical sensor mentioned above. It is described separately because, first of all it is a very popular solution, and secondly the mechanics do not hamper the user in the same way as the other types of mechanic sensors do. Optical fibres are placed along the limbs of a subject, e.g. the finger, and a signal is produced which reflects the bending of the fibre/limb. An example of a glove where this technology is used can be seen in figure 2.3.

Most of the above devices are mainly used in kinematic systems where the goal is to track spatial coordinates of segments or joints over time. The problems with them are that the subject whose motion is to be captured must be wired up, making it cumbersome to move around. This motivates the use of passive sensing.



Figure 2.3: A fibre-optic glove (the 5th Glove) for MoCap [3].

### 2.3.2 Passive Sensing

In passive sensing the idea is to use an image obtained from a camera<sup>1</sup> and capture the motion based on that image. The image is a 2D projection of the 3D world, so the trick is to find the 3D (human) motion that gave rise to the 2D projection. This problem is known as photogrammetric reconstruction [1]. The idea was first used by Muybridge back in 1887 where he set out to prove that a horse has a flight phase where all its limbs are in the air at the same time [23]. Later Muybridge turned his research to human motion capture. Recently, Bregler and Malik have successfully tried to use state of the art computer vision algorithms to track the original data used in Muybridge's early experiments on capturing human motion [7].

The idea of using passive sensing, e.g. a camera, to capture the motion of the subject is very novel but also very difficult. The difficulties arise because of the 3D to 2D projection and the amount of information in an image sequence. To reduce these problems many systems use markers attached to the subject, reducing the amount of information and making the photogrammetric reconstruction easier. Marey [15] used this idea by attaching white stripes between the main joints of the user - the first passive marker.

In 1895 Braune and Fischer were the first to explore the idea of using active markers<sup>2</sup> [6]. They attached thin light-tubes to different body segments and generated short bursts of light synchronously photographed by four cameras. This type of active marker has become known as Moving Light Displays (MLD), where an image sequence is reduced to a sequence of moving lights (see Cedras & Shah [9] for a review). A classical MLD example is the one done by Johansson back in the early 1970s [14]. He showed that human activities like walking and running can be recognised from lights attached to the joints of an actor. This suggests that pure motion can be used for direct recognition instead of doing it indirectly through a geometric

---

<sup>1</sup>More than one camera can be used.

<sup>2</sup>Even though the markers might be active, the overall sensing technology is still considered to be passive.

reconstruction. Today the idea of direct recognition is used by many researchers around the world (see e.g. Campbell & Bobick [8]).

Even though the use of markers is a good idea, it is still cumbersome for the user. Therefore computer vision researchers have recently tried to move away from the marker approach and aim at more "pure" MoCap systems, where the "raw" input image is used to perform the photogrammetric reconstruction (see Moeslund & Granum [20] for a review).

### 2.3.3 Complexity of Different Devices

When MoCap is used in practice the output from the MoCap sensor system (one or more devices) needs to be analysed before the motion data are available. This is illustrated in figure 2.4. The complexity of the analyser module depends on the sensor module. The higher the level of the data that the sensor module produces, the less complexity is required by the analyser module, and vice versa. For the devices based on active sensing the output from the sensor module is usually very high, yielding only a little complexity in the analyser module. For devices based on passive sensing usually high complexity is required by the analyser module<sup>3</sup>.

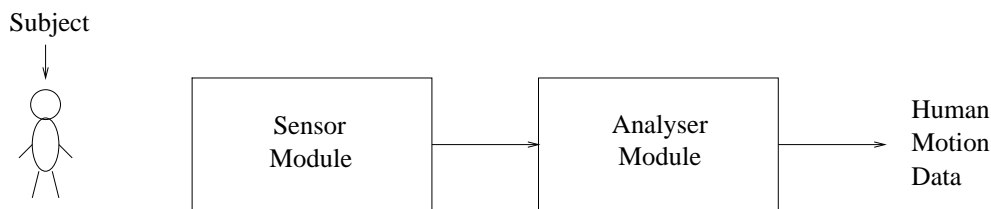


Figure 2.4: An illustration of the components in a MoCap system.

In MoCap systems based on active sensing the overall performance depends on the number of devices and their quality. The latest versions of these devices are fast (up to 140 Hz) and rather accurate (down to 0.5 mm in a well-controlled setup) [15]. If more devices are placed on the subject a better performance is obtained. In the analyser module not much performance can be gained when using active sensing, but the use of inverse kinematics may improve the overall result to some extent.

For passive devices, especially without markers, the performance is totally depending on the analyser module. The output from the sensor module does not contain any explicit information on the motion of the subject. The data needs to be analysed to extract the relevant motion data. A simple analyser module may only extract the motion of the silhouette of the subject, while a more complex analyser module may extract the actual pose of different limbs over time. In figure 2.5 the captured data, at one time instance, from two different systems using two different complexities, are shown. Figure 2.5.B shows the extracted silhouette of the left arm of the subject

<sup>3</sup>The analyser module becomes somewhat simpler when markers are used.

in figure 2.5.A. This is obviously the result of a somewhat simple analyser module, but can still be used in, for example, motion detection and texture mapping. In figure 2.5.C the 3D estimated pose of the left arm is superimposed on the input image, and it can be seen that the estimation is rather good. This is the result of a complex analyser module and is comparable to the devices based on active sensing, but without being as cumbersome.

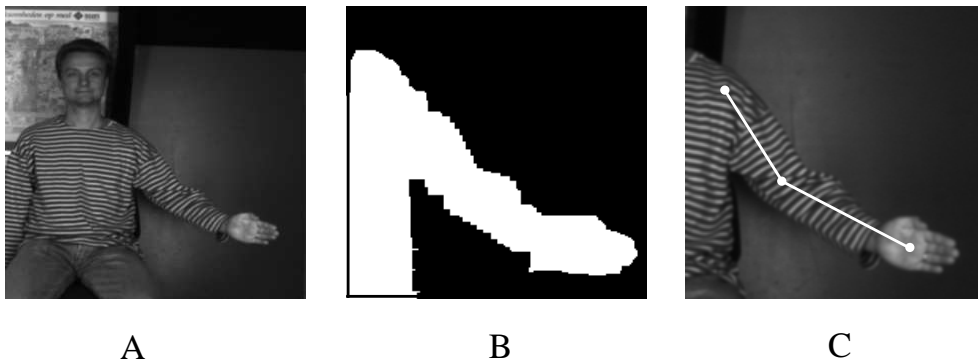


Figure 2.5: A: An input image. B: The estimated silhouette of the subject's left arm. C: The estimated 3D pose of a subject's left arm superimposed on the input image [19].

After having described the different MoCap technologies and devices, we will now see how these may be used in, particularly in control applications.

## 2.4 Motion Capture Used in Control Applications

MoCap have over the years been applied in many different applications. These may be divided into three overall areas concerning *surveillance*, *analysis*, and *control*, respectively.

The surveillance area covers applications where a subject or a number of subjects are being tracked over time and possibly monitored for special actions. A classic example is the surveillance of a parking lot, where a system tracks subjects to evaluate whether they are about to commit a crime, e.g. steal a car. Due to the nature of this area only passive sensing is applied in these applications.

The analysis area is concerned with the analysis of the raw motion data. This could be used in clinical studies of, for example, diagnostics of orthopedic patients, or to help athletes understand and improve their performance. Magnetic sensors and marker-based vision systems are the most common solutions in these applications. This is mainly due to the fact that the applications are situated in a highly controlled environment and due to the requirement of very precise motion data.

The control area relates to applications where the captured motion is used to control

something. It could be used as an interface to games, virtual worlds or animation, or to control remote located implements. Both active and passive sensing are used depending on the concrete application.

Since the interaction with virtual worlds is one of the main topics of this book we will look closer at the control area where this application is included.

After the motion of a subject is captured, it may, through an interface, be used to control something in either the real world or in a virtual world. Clearly the interface to the virtual world is more interesting in the context of this book, but we shall also say a few words about interacting with the real world.

When discussing these interfaces we only consider the control signals from the subject to the real/virtual world, i.e. a one-way interaction. Interface issues such as how the interactions may be visualised and by whom, and which data protocols are used, are not considered.

### 2.4.1 Interacting with the Real World Through a MoCap System

In figure 2.6 the interaction between a subject, through a MoCap device, and the real world is illustrated. The arrow indicates the interaction. One way of interpreting it, or the interaction, is to classify different interactions by the degrees of freedom (DoF) in the interface. Defining it like this, a device such as a button with one DoF will be located at one end of the axis, while a body suit or a full-blown computer vision system estimating all body parameters will be located at the other extreme.

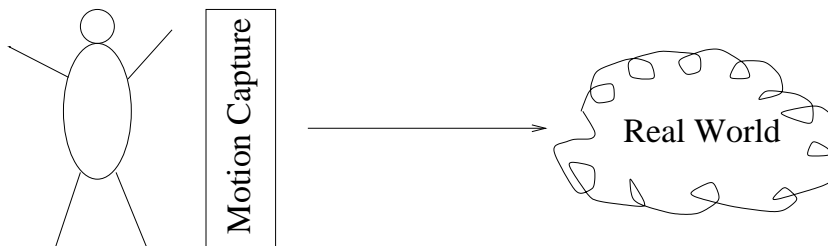


Figure 2.6: An interface from a subject to the real world through a MoCap system.

Another way of looking at the interaction is to divide interfaces into synchronous and asynchronous. This, we feel, will provide the best insight into different interaction methods.

Synchronous interaction is interaction carried out online. That is, the motion captured is immediately used to control something in the real/virtual world.

Asynchronous interaction is interaction concerned with recording. That is, when a subject's motion is captured it is not used immediately but rather stored for later use.

### Synchronous Interaction with the Real World

Synchronous interaction can be seen a lot using devices which are normally not considered MoCap devices, such as steering wheels, brakes, gears and buttons. The use of the devices which *are* considered to be MoCap devices (see section 2.3), is still being researched. In several movies a view into what might be tomorrow's synchronous interfaces to the real world can be seen. In the science fiction movie *Aliens* (1986) Ellen Ripley (played by Sigourney Weaver) successfully fights off one of the meanest aliens ever encountered on film using a so-called powerloader - a mechanical device structured like the human skeleton which Ripley steps into. Her movements control the arms and legs of the powerloader and thereby multiply her strength by a factor of several thousand. In the movie *Dave* (1993) the President of the USA (played by Kevin Kline) during a visit to a factory controls a giant robot using the same technique as in *Aliens*. Besides the size of the controlled robot, the difference is that the President is not inside the actual robot but instead controlling it from a distance. In *Lost in Space* (1998) a boy (played by Jack Johnson) controls a fighting robot from a distance, as in *Dave*, using a so-called "Holographic Interface". It is not evident how the motion is captured, but it is definitely based on passive sensing.

Even though these devices are still science fiction they might actually be implementable in a not too distant future.

### Asynchronous Interaction with the Real World

In many industries, especially where conveyer belts are used, a high number of repeated operations are seen. These are mainly carried out by machines/robots to insure similarity in the production and speed, and to avoid wearing humans down. As the operations become more complex it becomes more complicated to program them to carry out the operations. This problem has been tackled using asynchronous interaction. A human performs the operation, while a computer captures his or her motion. Afterwards the computer may use the captured motion data to control the same machine performing the operation. The best-known example is perhaps spray painting, where a human paints an object. The movements of the painting device are captured and can afterwards be replicated by a painting robot.

#### 2.4.2 Interacting with a Virtual World Through a MoCap System

Figure 2.7 shows a subject interfacing to a virtual world through a MoCap system. The arrow indicates the interaction between the subject and the virtual world. It may, as mentioned in Section 2.4.1, be representing the DoF in the interface. But, as before, we choose to use the synchronous/asynchronous division.

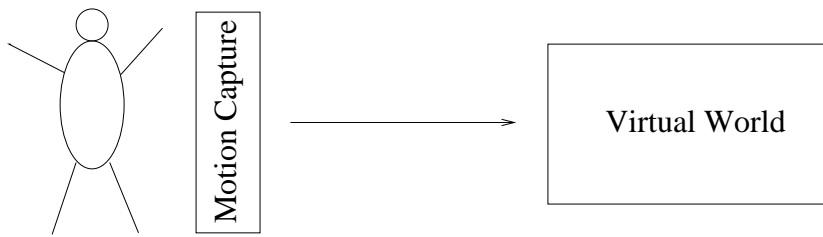


Figure 2.7: An interface from a subject to the virtual world through a MoCap system.

### Synchronous Interaction with a Virtual World

Synchronous interaction with a virtual world may be carried out in two different ways: either as a one-to-one mapping or as an interpretation. That is, a **direct** or **indirect mapping**. The direct mapping covers the methods where the subject's movements are duplicated directly in the virtual world. In the ultimate system this means that one could not tell the difference between the movements in the real world and the movements in the virtual world. Indirect mapping is when the subject's movements are captured, interpreted, and *then* used to control something. That is, the signals sent through the interface are discrete symbols rather than "raw" continuous motion data. The subject could wave his hand in the real world and it would be sent to the virtual world as a command, for example to close the connection between the two worlds.

In some systems both ways may be seen. For instance, in systems where the subject's movements are mapped directly into the virtual world there is the problem of moving around. The virtual world is usually bigger (in theory infinite) than the real world where the MoCap equipment applies. Therefore some method must be designed to allow subjects to move around in the virtual world without having to move (as far) in the real world. This can be done using a special mapping, e.g. walking on the spot maps to walking in the virtual world. Clearly this requires some sort of interpretation, but still the rest of the motions conducted by the subject may be one-to-one mapped, i.e. a mix of both methods. The two different types of interaction will be described further below.

When a **direct mapping** takes place it is sometimes known as **performance animation**<sup>4</sup> and the object animated in the virtual world is known as the **avatar** of the subject. A good example of performance animation is the work by Wilson where he defines performance animation as *animating a graphical character via a like human performance* [27], i.e. the animation is based on performed motion as opposed to, for example, using key-frames. In his Luxomatic system a computer vision system captures the motion of the subject's hand and uses it to control a lamp in the vir-

<sup>4</sup>Performance animation may also be based on prerecorded data, i.e. an asynchronous interaction.

tual world. Figure 2.8 shows a snapshot from the system. The MoCap process can be seen in the bottom-right corner, while the rest of the figure shows the subject's avatar<sup>5</sup>, the lamp, in the virtual world.

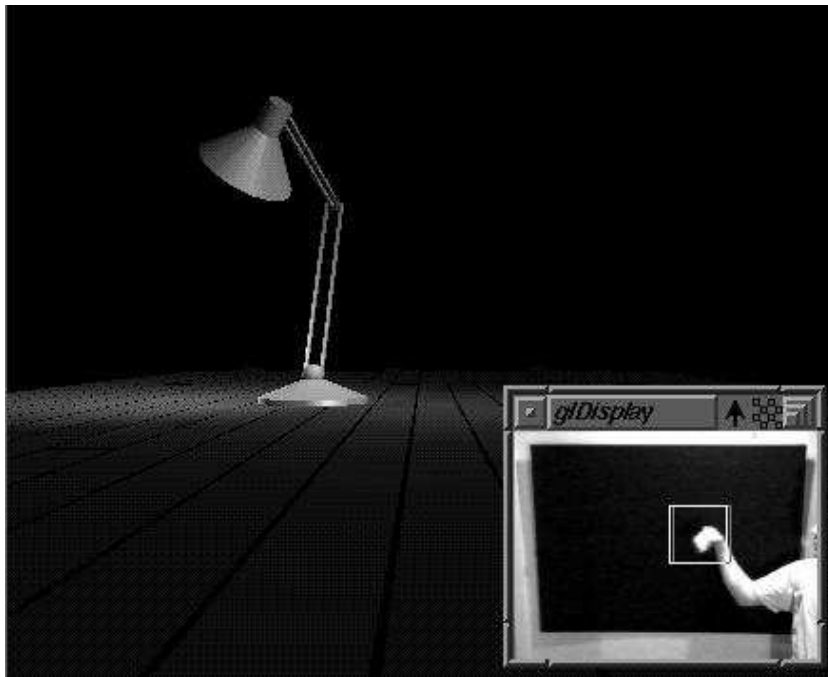


Figure 2.8: A snapshot from the *Luxomatic* system [27].

The term *performance animation* is mainly used when the animation is used as part of a performance, e.g. dance performance. That is, instead of viewing the performer the audience views the avatar. An example of this is shown in figure 2.9, where six images taken from a performance animated sequence, *Ghost*, made at The Department of Dramaturgy, Aarhus University, Denmark, is shown. The motion of a dancer is captured and mapped to the virtual character, which is animated as a fuzzy stick-figure.



Figure 2.9: A performance-animated sequence of a dancer.

This idea can be taken a step further where the performance is the *interaction* between the subject and his or her avatar! The term *performance animation* is

---

<sup>5</sup>Other good examples of avatars may be seen in movies such as *The Lawnmower Man* (1992) and *Disclosure* (1994).

rarely used in other applications even though it is a rather good definition.

In the case of **indirect mapping** the captured motion is not reproduced in the virtual world but rather interpreted and represented as a few symbols which are sent to the virtual world. In the work by Madsen and Granum [16] it is described how the motion of a subject's hand is captured by a dataglove and converted into one discrete symbol having one of three values (none/pointing/open hand). The symbol is sent to a virtual world where an autonomous agent uses it, together with an auditory input and a number of internal parameters, to decide how to react.

In the work by Freeman *et al.* [12] passive sensing without markers is used to capture the motion of a subject. The images are analysed and converted into a few symbols which explain the 2D orientation of the subject's hand. The symbols are used to control (steer) a car in a computer game.

### Asynchronous Interaction with a Virtual World

Good examples of where asynchronous interaction with a virtual world is used are the movie and advertising industries. Here MoCap data are recorded and stored, and later used to create animations. Animations have been used in cartoons in particular for a number of years. But animating complex objects moving realistically over time is somewhat difficult using hand animations, and therefore MoCap is used extensively, since it allows exactly this through performance animation. Imagine movies such as *Jurassic Park* (1993) and *Titanic* (1997) without computer animations based on MoCap. In the former, MoCap devices were placed on elephants and their movements were recorded. Later these motion patterns were applied to make the computer models of the dinosaurs come to life and move realistically. In the latter movie a number of people had their walking patterns captured and applied to human models to make them move realistically. Daily we actually also witness a number of computer animations based on MoCap. Look more closely next time you watch a computer graphic commercial on TV and wonder how they have made, for example, the small Colgate figure move so realistic - MoCap!

An entire industry specialising in MoCap for performance animation has emerged during the past five years. It consists of MoCap houses which all include a MoCap studio where high-tech equipment is used to capture motion for the movie and advertising industries in particular. A recent study found that about 30 to 40<sup>6</sup> MoCap houses provide a MoCap service for hire worldwide [10].

As mentioned above, this way of controlling objects in the virtual worlds is based on asynchronous processing. Of course, the MoCap process itself is carried out synchronous, but the data are afterwards stored and fine-tuned for later use. Therefore we consider this asynchronous interaction.

---

<sup>6</sup>Today the figure is more likely 100.

---

## 2.5 Discussion

MoCap is a powerful technology for designing advanced interfaces, especially when the interface is too complex for keyboard and mouse. The MoCap technology may be based on either active or passive sensing. Active sensing is generally based on wiring the subject up, while passive sensing is not. Obviously, this makes passive sensing most attractive in the context of this book.

Generally, active sensing gives rather precise data, but for many applications this solution is too cumbersome. Passive sensing may be based on either wearing or not wearing markers. When markers are used good results may be expected, but, as for active sensing, this is too cumbersome for some applications. When no markers are used a more general interface might be realised. The drawback is, however, that the complexity needed to extract and analyse the motion data is very high. Actually, no system has so far been able to capture the exact motion of an entire subject using only passive sensing without markers. This is therefore a large research topic and better and better solutions are being developed. In the future we might see workable systems using this technology.

MoCap may either be used in an synchronous or asynchronous interaction with virtual worlds. Synchronous interaction is interaction carried out on-line, i.e. the motion captured is immediately used to control something in the virtual world. When the motion data is not used immediately, but rather stored, fine-tuned and *then* used, the interaction is considered to be asynchronous. This is seen a lot, especially in big Hollywood movies and the advertising industry, and an entire industry has grown up to support it.

In the context of this book synchronous MoCap is most relevant. It is, as described earlier, divided into direct mapping and indirect mapping. When the interaction is used in a direct mapping it is known as performance animation and the object to be controlled in the virtual world is known as the avatar of the subject. Indirect mapping requires some sort of interpretation prior to use in a virtual world.

Whenever a human wants an avatar to mimic him or her the obvious choice is to use direct mapping. In all other situations, however, the motion data needs to be interpreted before an action is carried out in the virtual world, i.e. indirect mapping. Inspired by Andersen and Callesen's work [4] the interpretation can be viewed as a matter of converting the continuous motion data (trajectories) to discrete symbols according to some known sign system.

*Ad hoc* sign systems are being designed to suit different virtual worlds and the interpretation of these is an integral part of the MoCap process. Actually this simplifies the required analysis of the MoCap data, since it is easier to extract specific motion patterns than to capture general motion parameters.

However, if we wish to build good interfaces for people other than the designers we need to use the (both formal and informal) sign systems used in every-day life

situations. This means a lot of subtle signs which are extremely difficult to extract from the continuous flow of motion data. For example, consider a situation where a human moves his or her hand or head while saying: "They went that way". Clearly the movement indicates a direction, but unless the sign system is very limited or the context is known, it would be extremely difficult to convert the motion into a discrete symbol containing information on the indicated direction.

To reduce this problem the continuous motion data can be combined with the speech data and a multi-modal interpretation carried out. This idea is discussed further in the work by Paggio and Music [24].

## Acknowledgement

The author would like to thank Lone Koefoed Hansen (The Department of Multimedia, Aarhus University, Denmark) and Andrew D. Wilson (The Media lab, MIT, USA) for providing images for the Figures 2.2, 2.8 and 2.9.

## References

- [1] P. Allard, I.A.F. Stokes, and J.P. Blanchi, editors. *Three-Dimensional Analysis of Human Movement*. Human Kinetics, 1995.
- [2] K. Aminian *et al.* Motion Analysis in Clinical Practice Using Ambulatory Accelerometry. *Lecture Notes in Artificial Intelligence 1537. Modeling and Motion Capture Techniques for Virtual Environments*, 1998.
- [3] C.S. Andersen. A Survey of Gloves for Interaction with Virtual Worlds. Technical report, Laboratory of Image Analysis, Aalborg University, Denmark, August 1998.
- [4] P.B. Andersen and J. Callesen. Agents as Actors. In Lars Qvortrup, editor, *Interaction in Virtual Inhabited 3D Worlds*, chapter 7. Springer-Verlag, 2000.
- [5] N.J. Badler, M.J. Hollick, and J.P. Granieri. Real-Time Control of a Virtual Human Using Minimal Sensors. *A forum short paper in Presence, MIT*, 2(1):82–86, 1994.
- [6] C.W. Braune and O. Fischer. *The Human Gait*. Berlin: Springer-Verlag, 1987. Original work published in 1895.
- [7] C. Bregler and J. Malik. Tracking People with Twists and Exponential Maps. In *International Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, June 1998.

- 
- [8] L. Campbell and A. Bobick. Recognition of Human Body Motion Using Phase Space Constraints. In *International Conference on Computer Vision*, Cambridge, Massachusetts, 1995.
- [9] C. Cedras and M. Shah. A Survey of Motion Analysis from Moving Light Displays. In *Computer Vision and Pattern Recognition*, pages 214–221, Seattle, Washington, USA, June 20-24 1994.
- [10] B. Delaney. On the Trail of the Shadow Women: The Mystery of Motion Capture. *Computer Graphics and Applications*, 18(5):14–19, 1998.
- [11] S. Dyer, J. Martin, and J. Zulauf. Motion Capture White Paper. Technical report, Windlight Studios & Wavefront, December 1995.
- [12] W.T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma. Computer vision for computer games. In *International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, 1995.
- [13] W. Frey, M. Zyda, R. Mcghee, and B. Cockayne. Off-The-Shelf, Real-Time, Human Body Motion Capture for Synthetic Environments. Technical Report NPSCS-96-003, Computer Science Department, Naval Postgraduate School, Monterey, California, USA, June 1996.
- [14] G. Johansson. Visual Motion Perception. In *Scientific American*, pages 76–88. June 1975.
- [15] Z. Ladin. *Three-Dimensional Analysis of Human Movement*, chapter 1. Human Kinetics, 1995.
- [16] C.B. Madsen and E. Granum. Aspects of Interactive Autonomy and Perception. In Lars Qvortrup, editor, *Interaction in Virtual Inhabited 3D Worlds*, chapter 9. Springer-Verlag, 2000.
- [17] E.J. Marey. *Animal Mechanism: A Treatise on Terrestrial and Aerial Locomotion*. New York: Appleton, 1873. Republished as Vol. XI of the International Scientific Series.
- [18] T.B. Moeslund. Interacting with a Virtual World through Motion Capture. In Lars Qvortrup, editor, *Interaction in Virtual Inhabited 3D Worlds*, chapter 11. Springer-Verlag, 2000.
- [19] T.B. Moeslund and E. Granum. Multiple Cues used in Model-Based Human Motion Capture. In *The fourth International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000.
- [20] T.B. Moeslund and E. Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3), 2001.

- [21] T. Molet, R. Boulic, and D. Thalmann. A Real Time Anatomical Converter for Human Motion Capture. In *EUROGRAPHICS Int. Workshop on Computer Animation and Simulation*, Poitier, France, 1996.
- [22] T. Molet, Z. Huang, R. Boulic, and D. Thalmann. An Animation Interface Designed for Motion Capture. In *Conf. on Computer Animation*, 1997.
- [23] E. Muybridge. Animal locomotion, 1957. Reprinted in Brown, L.S.(Ed.)(1957). *Animal in motion*. New York: Dover.
- [24] P. Paggio and B. Music. Linguistic Interaction in Staging - a Language Engineering View. In Lars Qvortrup, editor, *Interaction in Virtual Inhabited 3D Worlds*, chapter 12. Springer-Verlag, 2000.
- [25] Lars Qvortrup, editor. *Interaction in Virtual Inhabited 3D Worlds*. Springer-Verlag, 2000.
- [26] S.K. Semwal, R. Hightower, and S. Stansfield. Closed Form Geometric Algorithms for Real-Time Control of an Avatar. In *International Symposium on Virtual Reality*, Santa Clare, California, USA, 1996.
- [27] A.D. Wilson. Luxomatic: Computer Vision for Puppeteering. Technical Report 512, MIT Media Lab, 1999.
- [28] C. Youngblut, R.E. Johnson, S.H. Nash, R.A. Wienclaw, and C.A. Will. Review of Virtual Environment Interface Technology. Technical Report H 96-001239, Institute of Defense Analyses, 1996.