Abstract. In this paper we provide the reader with a simple, yet thorough, overview of what is motion capture, its history and evolution so far and processes used in acquiring the tri-dimensional data of the recorded scene, take or animation. A brief look at the “mocap” process’s (short for motion capture) structure is given followed by an extensive list of the technologies behind the most popular systems used today. We dive into the inner works of acoustical, mechanical, magnetic and optical mocap systems, also discussing the differences between active, passive and markerless optical systems, since these are the most popular of the above referred. Later on we also provide some insight into facial motion capture, right after we compare the data acquisition systems and quickly overview the generic data file structure. Finally we provide examples of real-world applications and some possible research fields in the area along with our conclusions.

Keywords: Motion capture, movement reconstruction, markerless capture.

1 Introduction

Animation gave its first steps in the early 20th century, when in 1911, cartoonist Winsor McCay drew a character in multiple sheets of paper with slight changes between these and then sampled them at a constant rate to create the illusion of motion [1].

Animation processes did not witness considerable innovation until computers started to take place in the process. With the birth of keyframing, which reduced the amount of samples needed to create an animation animators saw their work a lot more simplified. This process was time consuming because, at the time, every artist was forced to individually animate each pose/frame. With the introduction of keyframing the artist specified the initial and ending frames of the animation and the intermediate frames of the movement where automatically generated [2]. However some animations were still impossible to recreate due to their inherent complexity, for example the human walking animation, which is terrifyingly complex due to our articulations.

To speed up the animation process further, motion capture was invented, a
means by which we capture the movements of objects in the real world and then insert the data of the captured movement in a tridimensional model of the world in a virtual environment. The process first evolved with mechanical systems that were quite cumbersome and limited the amount of freedom the actor could experience, limiting severely the animation spectrum that could be captured. This happened mainly because these were mechanical systems that resorted to very restrictive suits and large amounts of cable that hindered the actor’s movements. Mocap has evolved much since then and today, as a response to these early issues, we have a broader range of options (Section 3). They include acoustical, mechanical, optical and magnetic systems, further divided in marker and markerless systems.

Today, motion capture is widely used in the gaming, movie and animation industry [3] as a means to provide quick, budget adapting body and/or facial animations in order to animate one or various characters. We provide insight into these methods and processes, and also the data processing and data formats that most systems use. Lastly we introduce some future work and research in motion capture. Research we believe would be highly beneficial and would enable future developments and breakthroughs in the area.

2 State of the Art

Motion capture has come a long way since its birth a few decades in the past. Still, there are a number of pending issues waiting a solution. In this section we provide some systems that represent today’s state of the art in motion capture and that have dealt and overcome some of the limitations of mocap. We start by introducing, Ascension’s Motion Star system [4]. This system tries to overcome some of the shortcomings of magnetic mocap systems, such as degree of freedom offered and electro-magnetic interference. Stanford University has also developed state of the art systems for markerless motion capture [5], that provides great accuracy, full body calculation of joint angles and joint centres, as well as a model matching followed by kinematic extraction. Carnegie Mellon University has also developed similar systems [6], which provide fully articulated 3D models by means of a joint skeleton and body shape acquisition. Finally Image Metrics has developed a truly astonishing markerless motion capture system, based in a spherical lighting rig equipped high resolution digital cameras and their proprietary markerless mocap technology that captures the model’s animations to a degree where it is almost impossible to differ between the real and the virtual actor. We refer and council the reader to witness this work in [7] and [8], for this system is truly a noteworthy feat.
3 Motion Capture Systems

3.1 Marker-based Motion Capture

3.1.1 Acoustical Systems
In this type of system a set of sound transmitters are placed on the actor’s main articulations, while three receptors are positioned in the capture site. The emitters are then sequentially activated, producing a characteristic set of frequencies, that the receptors pick up and use to calculate the emitters’ positions in three-dimensional space. The computation of the position of each transmitter is as follows: Using as data the time interval between the emitting of the noise by the transmitter, the reception of this one by the receptor and the travelling speed of sound in the environment, one can calculate the distance travelled by the noise. To determine the 3D position of each transmitter, a triangulation of the distances between the emitter and each of the receptors is computed [9].

Some problems with these systems are; firstly, the difficulty in obtaining a correct description of the data in a certain instant. This is due to the sequential firing nature of the transmitters, which can create a non-fluid, or not so fluid as desirable, description of the movement. Another downside of these methods are the restrictions to the freedom of movement the cables induce on the actor, reducing the scope of movements available. Finally, the amount of transmitters that can be used is also limited, a factor that can hinder the quality of the animation [9]. Although these systems do not suffer of occluding or metallic object interference issues, typical in optical and magnetic systems, they are very susceptible to sound reflections or external noise, which being quite frequent and uncontrollable in most situations make them a "last resource” choice.

3.1.2 Mechanical Systems
These systems are made out of potentiometers and sliders that are put in the desired articulations and enable the display of their positions (Figure 1). Despite being underdeveloped, mechanical motion capture systems have some advantages that make them quite attractive. One advantage of these systems is that they possess an interface that is similar to stop-motion systems that are very popular and used in the film industry, thus permitting an easy transition between the two technologies. A final advantage is that they’re not affected by magnetic fields or unwanted reflections, not needing a long recalibration process, which makes their use easy and productive.
3.1.3 Magnetic Systems

Using a set of receptors that are placed in the actor’s articulations it’s possible to measure the positioning and orientation of the articulations relative to an antenna. Magnetic systems aren’t very expensive, in comparison with other systems for motion capture. The workstation used for data acquisition and processing is cheap as well and the precision of data is quite high. With a typical sampling rate of \( \sim 100 \) frames per second, magnetic systems are perfect for simple movement capture.

The disadvantages of these systems include the huge number of cables that connect to the antenna, reducing the freedom degrees of the actor. Systems that do not require the use of cables are currently being developed [10], effectively eliminating this drawback. The possible interference in the magnetic field caused by various metallic objects and structures poses a restriction to the surrounding material, which can be of some gravity. Some systems are highly sensible even to the building’s own structure presenting some interference, making this a critical flaw in magnetic mocap systems.

Nowadays we observe an effort by these product’s manufacturers, which are investing in new and improved versions of their systems that exhibit a reduced vulnerability to these issues. Nonetheless, the issue has not been completely eradicated. Further reading can be done in [11].

3.1.4 Optical Systems

With this kind of system the actor wears an especially designed suit, covered with reflectors that are placed in their main articulations. Then, high-resolution cameras are strategically positioned to track those reflectors during the actor’s
movement. Each camera generates the 2D coordinates for each reflector, obtained via a segmentation step. Proprietary software is then used to analyse the data captured by all of the cameras to compute the 3D coordinates of the reflectors. These systems are the most expensive ones in the market due to their cutting-end technological nature, such as the high-resolution cameras and sophisticated proprietary software. The values can hit the USD 250,000 bar.

The advantages of using these systems are mainly the very high sampling rate, which enables the capture of fast movements such as martial arts, acrobatics and gymnastics, among others. The sampling rate usually depends on the cameras used, meaning that the higher the resolution, the higher the sampling rate will be. Sampling rates of up to 200 frames/second are achievable.

Another advantage is the freedom offered by these systems, since, unlike the other systems, there are no cables or limited workspace and the reflectors pose no restrain or cumbersome effect on the actor. Also, since the reflectors offer no resistance there is virtually no limit to the number used in the capture process, which enables a very high theoretical level of detail.

The disadvantages of this method are the occlusion of some transmitters, especially in small objects such as hands or closely interacting objects, a problem that can compromise the entire process if the occluded data is unrecoverable. This problem can be somewhat overcome with the deployment of additional cameras and/or reflectors but this implies a higher processing time for the CPU during the tracking. Increasing the reflector numbers is no panacea, since the "tracking-confusion" problem will eventually arise. This issue consists on the difficulty in identifying reflectors that are too close to one another and is directly influenced by the resolution of the cameras, being thus solved, by using cameras with higher resolution capabilities. This obviously leads to budget issues as these cameras are, as previously explained, extremely expensive. One final problem we can encounter is the lack of interactivity, since the data gathered must be processed (and sometimes undergo filtering and noise reduction) before it is usable. This gives us little feedback, an inconvenience when production costs are involved, the studio is booked for only a few hours and time is of the essence. Further reading on these systems can be done in [12].

Figure 3: Upper left: A mocap studio. Upper right: Same mocap studio during capture. Lower left: A motion capture camera. Lower right: Close-up of a motion capture camera.

Active Markers

This kind of optical motion captures use LED’s that, instead of reflecting the light emitted by the high-resolution cameras, emit their own light, being powered by a
small battery. Since according to the Inverse Square Law [13] we can achieve 1/4 the power at two times the distance, this can increase the usable capture volume.

**Passive Markers**

These markers, unlike the active markers, are coated with a retro-reflective material that reflects the light back to the cameras, that must first be calibrated so only the markers are identified, ignoring other materials. Using multiple cameras they are calibrated using an object with reflectors for which the positions of these are known. Waving a “wand” imbued with a series of reflectors, across the capture volume, is how this calibration process is usually performed. Normally a system will incorporate anywhere from 6 to 24 cameras, but some systems with over 300 cameras exist to reduce marker swap or confusion issues in complex captures.

![Figure 4: Passive markers placed on an actor’s face.](image)

**3.2 Markerless Motion Capture**

The ever-growing research in computer vision fields is quickly enabling the development of markerless optical motion capture techniques. These systems do not require special equipment for tracking of the actors’ movement. The actors’ movement is recorded in multiple video streams and computer vision algorithms analyse these streams to identify human forms, decomposing these into single, isolated parts used for tracking. The motion capture process is, thus, completely done via software, removing all the physical limitations while introducing computational constrains. A perfect example of such a system is Microsoft’s Kinect, which has successfully introduced a solution for low-cost motion capture to the masses [14]. We refer the reader to [15], [16] and [17] for further reading on some of these systems.

![Figure 5: Markerless based motion capture used in the Lord of the Rings trilogy: Upper image: The actor whose movements are going to be captured. Lower image: The computer generated character side-to-side with the actor imitating his movements.](image)
4 Acquired Data Processing

In what regards the data acquisition type, mocap systems are classified in two categories; direct and indirect acquisition. Direct acquisition systems are the ones that do not require any type of post-processing cycle, such as magnetic, acoustical and mechanical systems. Although this is, in fact, an upside to these systems it does not come without drawbacks. This is because these systems are more obtrusive and offer a lower sampling rate [18].

Indirect acquisition systems, in which optical systems are included, offer more freedom to the performer and a higher sampling rate. These systems group the captured data in data sets that are processed \textit{a posteriori} by dedicated software; responsible for generating the final data set that represents the captured movement. This workflow scheme makes the use of these systems impossible in projects that require interactivity and near instantaneous viewing of the captures. They are especially suited for the accurate and seamlessly perfect capture of complex and fast movements [18].

Although not all of the methods described earlier require a data processing cycle, all of them, require the data to be cleaned, filtered and then mapped (rigged) to a skeleton, which then follows the positions of the markers in every frame of the animation by binding the markers to the skeleton’s own articulations.

4.1 Data Formats

Data formats are normally proprietary formats, owned by the company that developed the software. These usually follow the same structure, only differing in some minor nuances like commentaries, special symbols, data and time information and additional data that the manufacturer deems useful.

The body parts are segmented into key regions and the information for each segment is stored separately. The segment structure consists of an object vector, with each vector representing one sample of the segment, stored in some file format (e.g. 3DS). Every segment object is represented by a 3D model with a bounding box and contain 9 information channels relevant to the segments X, Y and Z positions, as well as orientation and scale factors for each coordinate. This data structure makes it possible to transfer the animations between different models, given that they have the same scale and pivot positions in every segment [18]. There is research in this field to create algorithmic procedures that adapt the data to other structures without producing unrealistic or faulty animations. This would be ideally done by adjusting the pivot positions of the segments, the scale factors and the weights associated within each segment in the skeleton of the mode.

5 Applications

5.1 Films

5.1.1 King Kong

The mocap techniques used in the animation of the King Kong creature in Peter Jackson’s film adaptation of the novel with the same name, involved optical marker based mocap with around 70 to 80 cameras in a huge warehouse, adapted for this function. Every animation of King Kong was played out by one actor (Andy Serkis)
and then later heavily edited to compensate for the physiological differences between
him and the skyscraper-climbing ape. In the capture about 60 optical markers on the
body and 12 on the face were used. Out of the 70 to 80 cameras, 52 where used to
track the body movement and around 20 for the facial expressions, the number
however varied slightly from scene to scene [19].

The greatest challenge, however, came with the fact that the producers didn’t
want human-like movement, but ape-like movement. So, even though, the actor was
acting like an ape, this just couldn’t compensate for the fact that his body wasn’t that
of an ape, much less that of a five-story building one. Having taken that into
consideration, a specific piece of software was used to translate the movement of the
actor into a more convincing one, compensating for all the physical discrepancies.
The software mapped every marker in the actor’s body to a corresponding point in the
gorilla’s body and then adjusted the movement as it was being recorded. After this
phase, the animation still went through minor adjustments done by some technicians
to re-touch and polish some nuances [19]. Unfortunately the software used was
proprietary, so no more details other than this basic workflow specification were
released to the general public.

5.1.2 Lord of the Rings: The Return of the King

The visual effects of the Lord of the Rings trilogy were put in charge of WETA
Digital [20], which assembled a team of computer artists, motion editors, and
software engineers, among others. They also created a never-before-seen database to
store every single shot frame in the trilogy. This digital library was able to track,
cross-reference and analyse every item appearing in the film, making every single
element susceptible to digital manipulation. This included characters, animals, props,
and so on, meaning that the motion capture process had no need for extra assets,
which reduced occlusion and movement restriction issues introduced by these items.
Such an approach to these problems was a novelty in the area.

Despite this, the most astonishing achievement by WETA was the creation of
fully digital creatures, featured in the game, such as Gollum (Figure 5), Treebeard, the
Balrog and the Eye of Sauron. Gollum is one of the most surprising achievements in
the mocap process of the Lord of the Rings trilogy as it was created by a mixture of
motion capture technology and state-of-the-art computer animation. Also, in order to
distance themselves from normal computer-generated characters Gollum was
animated using joint movement based on organic muscle and bones, which can be
seen in action in the film as the muscles and bone move under his translucent skin
[21].

Citation: "WETA developed vast amounts of code to create Gollum", "They
developed new modelling codes, new skin codes, new muscle codes. He is amazingly
life-like and we were able to give him a range of expressions from the evil of Gollum
to the sympathy of Smeagol.” - Peter Jackson in LotR Making Of.

Optical motion capture techniques were used to record the actors and proprietary
software to animate the creatures and battle scenes. These animations created via
software were constantly subjected to synchronization with the mocap actor’s
movements (and also between the digital creatures), in order to create a believable
battle or dialog scene [22].
5.2 Videogames
5.2.1 Tiger Woods PGA Tour 07
Tiger Woods used a specifically developed facial motion capture technique, called Universal Capture, by Digital Air [24]. The process starts with a head scan, through the use of high-definition cameras, to obtain a 3D model of the actor’s head. Then a large amount of markers (from 80 to 90) are placed in key regions of the actor’s head and the movements of the markers are recorded as animations in a three-dimensional space by the cameras. The movement, opposed to FIFA’s, which was captured in a spacious room, is recorded in a confined booth with limited movement. This is justified by the fact that in this process we only want facial animation and not full-body animation, where the players show virtually no facial animation. About 18 mocap cameras, 3 video cameras, 2 microphones and a huge amount of floodlights surround the booth. The three video cameras are used to capture the actor’s face in three different angles, and the recorded video is then used to wrap around the modeled face from phase 1 of the capture process and thus create a final animated 3D model of the actor’s head, which is animated via the recorded data by the mocap cameras and markers, in the booth [24]. A video demonstration and walkthrough of the system can be seen in [25].

5.2.2 Lord of the Rings: The Return of the King
The Return of the King used the same mocap systems (marker-based optical motion capture) that the production crew of the film trilogy choose. They used available animations, borrowed from the films and even the same actors that participated in the film were recruited. This was done in order to achieve maximum synchronization between the original animations and the game’s animations, such as the combat or cutscenes’ movements. We can therefore conclude that the mocap process from the films didn’t need to be adapted because the new animations to be recorded were doable in the already used system. This was because, being an optical mocap system, gave the actors enough liberty to act out the desired animations.

6 Conclusions
Mocap systems, as shown throughout the paper, have evolved from simple, highly restricting, user un-friendly systems (software wise), to very mobile and specialized ones. The types of system discussed clearly all have their optimal case scenarios for deployment. However, optical mocap has evolved much more than its brethren systems. This is, for the most part, due to its ability to adapt very well to the major requirements of the film and videogame industry, which have invested and thus aided in this technology’s development.

Despite its advantages, there are still improvements that can be done in this field (some of which proposed in the Future Research section). Other than these improvements, some mocap systems are still very limited in terms of the area where they can capture movements, being restricted to adapted warehouses or studios. These systems are also very high budget, which, in some cases, rules them out of question. They could benefit from lower budget versions, more accessible by the public and smaller companies.
Synthesizing, acoustical mocap systems are a last resort due to their cumbersome, external sound or sound reflection inflicted noise and non-fluid results. Mechanical systems are best suited for movements that don’t require a high degree of freedom and there may be magnetic or reflection as these cause no noise in the capture process. Magnetic systems are useful in situations where the level of detail needed is high and the budget is low. However their great sensibility to the surrounding magnetic fields proves to be a critical flaw. We assumed the use a non-wired magnetic suit; otherwise, the captured movement’s choices would be reduced to low level of freedom ones. Optical systems, although being expensive and having a processing/noise reduction stage, prove to be the best solution in most of the cases. This is because they offer a very high level of detail, freedom and working area enabling the sampling of virtually any movement.

In conclusion, mocap systems are very useful tools and provide a solution to a very important problem as well as enable truly amazing effects, as shown in the article. We have introduced the basic methods behind the various mocap systems and explored their advantages and disadvantages, concluding that every system has its best use in a specific scenario (and budget). The reader was also introduced to some of this technology’s uses in real life and research topics, which we hope has provided a comprehensive overview of the area.

6.1 Future Research

After thoroughly studying the area and writing this paper we believe the following subjects would prove interesting fields of research.

Captured Movement Modification:
Since every data file represents a limited and closed data set of animations these cannot be manipulated after the caption process. An interesting field of research would be how to modify or derive new movements from a already captured movement.

Captured Movement Fusion and Concatenation:
As an extension to the prior research field one could try to create a fusion of any number of movements or concatenate them in order to create new movements from these original ones, while maintaining their fluidity and naturality, through algorithmic approaches.

Improvements in Actual Tracking Techniques:
As previously discussed, marker occlusion in optical mocap systems is critical. Improved tracking techniques could be developed to eliminate this problem.

Marker Mapping Techniques for Non-Human Beings:
Sometimes we want to capture movements that don’t belong to human beings. Studying how to position the markers in the most effective way in these beings could prove beneficial, especially if the procedure could adapt itself to every object, thus becoming general and optimal and if it could help solve the problem stated in the above suggestion for future research.

Markerless Mocap in the Creation of High Definition Realistic Animations:
A highly advanced system such as the one in [3] and [4] constitutes an important milestone in highly realistic animation. We find it a very interesting and promising research field, which can help new breakthroughs in motion capture. Would it be in the development of even more realistic games, movies featuring indistinguishable
virtual actors, personalized helpdesks, AIs with a human interface that creates credible and high definition expressions, amongst others.

**Higher interactivity in mocap systems that involve a pre-processing stage:**
With more interactive systems we would overcome the issues present in the systems that employ a data processing stage, enabling their use in projects that require nearly immediate access and validation of the gathered data.

**Transference of motion captured animations between models:**
Although mocap is somewhat transferable between models there are still a lot of limitations as to which animations can be transferred, given the source and destination model. For example, the animation of a normal person cannot be transferred to a giant, since by physical laws they move in a very different manner, or between models with very different body constitutions. An interesting case study would be how to adapt certain parameters of the animation so that it would be applicable to other models while maintaining its credibility.

**Markerless Motion Capture for Psychological Analysis:**
With the development of markerless mocap we could research methods to analyze an individuals inner feelings through expressions for psychological research and treatment. One application example of this technology would be an interactive tool that would respond, via a programmed AI, in accordance with an autistic child or mentally handicapped individual in order to further the healing process, since a human being while qualified cannot evaluate the person’s mental status as accurately or quickly as a well programmed computer.

**References**


