Computer Animation and Games

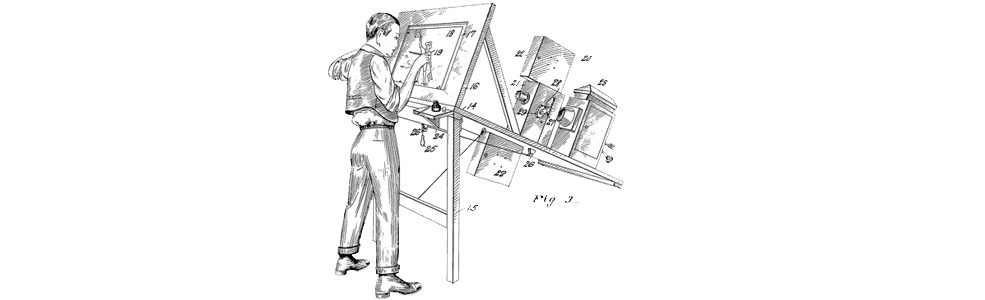
Development Overview of Motion Capture: Past, Present and possible directions in the Future

0. Abstract

Motion capture is a crucial tool for deploying natural movement into computer-generated animation both for animated films and for games, containing imaginary characters. The process of efficiently recording 3D positional coordinates for each limb either of a human or an animal, was researched even prior to the release of the so called personal computers (PC), but during the last couple of decades motion capture has reached its full potential producing state of the art character movement. This survey’s sole purpose is to give an overview of the history of motion capture as well as providing a brief but yet substantial explanation of the contents of various research papers containing any information about motion capture.

1. Introduction

Animation is the concept governing many aspects of today’s Digital Entertainment sector, with applications in computer animation, game development and of course visual effects. The notion of animation opened thousands of opportunities for every sector mentioned, since without the process of animating living or even inanimate objects, what is shown on the screen would not move. Motion capture (mocap), even though considerably argued, is responsible for taking animation to the next level, by giving imaginary characters realistic and natural motions. Mocap was introduced well before computers ever existed, with the first steps done completely by hand, calling the procedure “rotoscoping”, (Maher, 2017). Max Fleischer, 1914, came up with the idea of rotoscoping which is actually tracing over film footage of actors, frame by frame, where each paper represented a new frame, as can be seen in Figure 1. This technique has first surfaced by Disney Studios in their feature animated film “Snow White and the seven dwarfs”, in 1937. As one might have understood, rotoscoping is a timely procedure, but with the birth of the computing world, mocap has come a long way since then. This survey will explore the span of technology adaptations of motion capture from the beginning of the computer-generated animation era until now while separating the different techniques into categories and explaining their major differences. Mocap technologies can be distinguished into Optical Systems and Non-Optical Systems, which can then be separated, further, into other sub-categories.



2. Optical Systems

2.1 Marker- Based

2.1.1 Passive Markers

2.1.2 Active Markers

2.2 Markerless

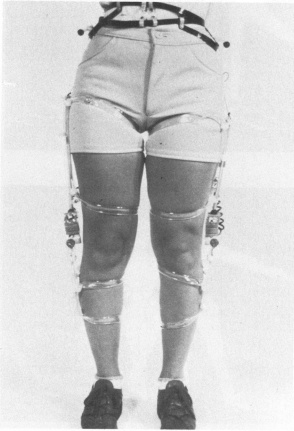
3. Non-Optical Systems

Motion capture is not restricted to visual data capturing, form images or videos. As a matter of fact, technologies that helped shape the notion of mocap as we know it, consisted of non-optical systems. There are various mechanisms that can be used for non-optical mocap each with a different apparatus approach.

3.1 Mechanical Motion

Mechanical motion capture concerns the recording of limb position using a mechanical suit or by controlling the orientation and shape of a mechanical arm, eventually mimicking the articulation of the joints of a character. Mechanical mocap was focused mainly on real-time applications, applications that could provide translation of joint movement directly to computer generated characters for advertising campaigns during major gaming and animation conventions. Real-time capabilities for mechanical mocap arose from the fact that data acquired did not require further post-processing to be used.

3.1.1 Exoskeletons

Motion capture suits, or in this case exoskeletons, were initially consisting of a series of electrogoniometers, apparatus using resistance to determine joint angles, attached to the bendable joints of a human body, as imagined and implemented by a group of computer scientists at the Simon Fraser University, (Calvert, Chapman and Patla, 1982), Figure 3.1. Once a joint, say elbow, was bent, the electrogoniometers would bend at the same time, in the same speed, thus recording natural movement passed as output, which could then be used to drive computer generated characters, producing animation sequences. Nonetheless, the exoskeleton was unable to track the real-world coordinates of the limbs, as they could only be used to capture certain movements, e.g. bends. The study focused on choreographic movements as well as assessing motion abnormalities, and even consisted of integration of the method mentioned, direct input from instrumentation, with notation from movement notation systems, indirect input, such as Kinetography Laban, a type of dance notation. By integrating both mechanisms, Calvert et al. were able to incorporate the output in such a way that could also be used for future reference, producing more life-like animation.

Since then a lot of companies tried to reproduce their own implementation of exoskeleton motion capture techniques, all following the structure of the implementation mentioned above. Mocap focusing companies started calling these types of suits as Waldo devices. In 1988, waldos became the, what thought at the time, next big development in mocap technologies. Specifically, computer animation giants Pacific Data Images, looked to employ a team that could produce an accurate waldo suit for their upcoming animations feature films. After two tries, Rick Lazzarini’s company, The Creature Shop, were able to produce an upper body and head exoskeleton based on optical potentiometers on each joint, (Menache, 2000). Carl Rosendahl, a part of the team working on the waldo, described the difficulty during production due to the noisy output of the analogue parts of the suit, leading to the replacement of analogue components with digital counterparts. The suit was used in the first ever motion picture that successfully incorporated a digital character called Toys, starring Robin Williams.

As years passed, mechanical mocap helmets entered the race for most natural motion in animation and gaming facial expressions. SimGraphics teamed up with Nintendo to produce an interactive Super Mario to showcase at the SCES convention in 1992. They were able to present a real-time floating Super Mario head able to interact directly with an audience at the 1992 SCES convention, (Silicon Graphics and Nintendo, 1992). The major development step taken to produce this new technique was the construction of a helmet, called the “Face Waldo”. The helmet consisted of various mechanical sensors attached to the primary facial parts that could reproduce human-like facial expressions, in turn portraying perfectly human emotions. These sensors were placed on the chin, lips, cheeks and eyebrows of the actor, that eventually controlled Mario’s corresponding digital facial parts. These types of sensors were used in concert with other electro-magnetic sensors, responsible for tracking the facial parts not covered by the mechanical sensors, to gather the much-anticipated input consisting of real-world coordinates of the facial parts of the actor. The input was directly mapped to Mario’s face and together with the “flying mouse”, a mouse that when moved in the air would control the position of Mario’s face on the workstation, and the voice of the actor, the famous cartoon and video game character came to life.

The most advanced mechanical mocap currently available is the Dexmo glove by Dexta Robotics. Mechanical mocap forced itself into the Virtual Reality (VR) industry due to the real-time aspect it could provide. The VR industry was in need of such a technology since it is required to wear a Head Mounted Display (HMU) in order to enter the VR world, which occludes the visual field of the user, specifically the user’s limbs. In order to avoid this, VR companies utilised the notion of mocap to allow users to see their limbs, e.g. hands and legs. On the plus side, other than using the exoskeletons for mocap, companies took advantage of this by utilising a force feedback mechanism in order for users to interact with virtual objects.

3.1.2 Digital Puppetry

With increasing interest in real-time motion capturing and animation rendering during the late 1980s, the term digital puppetry became the centre of attention. Digital puppetry is the process of controlling any 2D or 3D computer-generated character’s movements and other actions such as speech, in real-time with the use of computers. During a SIGGRAPH convention in 1988, “Mike, the Talking Head”, a digital character, played a leading role in the first ever digital live interactive performance, (Menache, 2000). Silicon Graphics was keen on showcasing their real-time processing and rendering of their so called “4D models”, with the fourth dimension being time, where they employed the likes of Wahrman and deGraf to produce an interface in which the input from a puppeteer device, the device that resembles a glove to move the lips and produce the “mouthing” motion of lips during the exclamation of words, was directly connected to the rendering engine of the polygonal character, Mike, and was able to interpolate frame instances between each motion. This was reproduced in front of an audience, in real-time in order for Silicon Graphics to legitimately validate the processing capabilities of their recently upgraded 4D workstation. The face mesh was produced using a 3D digitizer to scan the face of Mike Gribble, thus the name of the character, and consisted of over 200000 digital data points. With a speech recognition system and the glove mechanism, as well as scans of the real Mike as he mouthed certain phonemes, the lot were able to give digital Mike a personality. The notion of digital puppetry governed the mocap development process with other projects such as Waldo C. Graphic, another digital character controlled and animated in real-time using again the Silicon Graphics 4D workstation. Jim Henson Productions and Pacific Data Images were able to construct a mechanical arm with eight degrees of freedom, to control the position of the character and its movements in the screen, and a two-piece oven-mitt kind of mechanism attached at the very end of the mechanical structure, to control the lips of Waldo C. Graphic, (Jim Henson Productions, 1988), as can be seen in Figure 5.

3.2 Inertial

Inertial mocap systems are considered crucial in the development of gaming controllers, even though their history is not of the same length as the other systems mentioned in this survey. Inertial systems usually consist of inertial sensors and Inertial Measuring Units (IMU), which in turn can track rotational rates and local coordinate positioning using gyroscopes, accelerometers and magnetometers. As sensors and electrical components such as the gyroscopes or accelerometers, became smaller and cheaper over the last couple of decades, the inertial mocap system was made possible. As explained in the section of marker-based mocap, the more IMU sensors on the movable object, the more accurate and natural the data acquired is. This systems have the disadvantage of suffering from inertial drift, given an magnetic interference from common objects, which in turn ruin the accuracy of results and producing a jittering effect in the data collected.

3.3 Magnetic