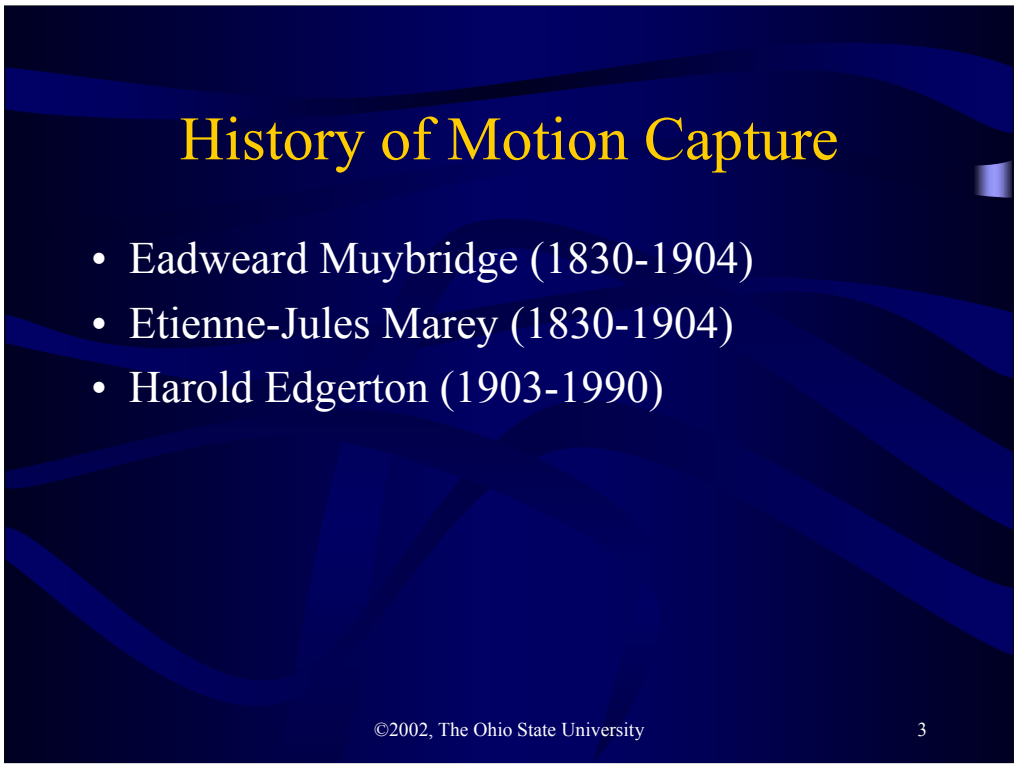
Motion Capture History and Pipeline

Suba Varadarajan

Advanced Computing Center for the Arts and Design
Ohio State University
varadara@cgrg.ohio-state.edu
http://www.accad.ohio-state.edu/~varadara

Motion Capture

- "Recording of motion for immediate or delayed analysis and playback"
  - David J. Sturman [1]



Marey and Muybridge conducted human and animal motion studies by shooting multiple photographs of moving subjects over a short period of time. Their work had a large impact on many disciplines such as biology, medicine, photography, and animation. Harold Edgerton was the inventor of flash photography.

Eadward Muybridge

- The Flying Horse
- Zoopraxiscope
- Animal Locomotion

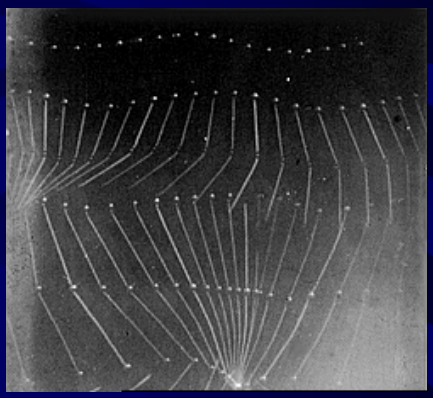
© Kingston Museum

Edward Muybridge first became involved in the photography of movement in 1872 when he was asked by Leland Stanford, the former Governor of California, to photograph his horse Occident. At the time, it was not known if a horse ever had all four feet off the ground while trotting. Muybridge was able to prove this true in his initial studies. Stanford subsequently financed a more elaborate investigation at Palo Alto from 1878-1879 where cameras were placed in a line fitted with a special shutter that could be triggered electro-magnetically by the horse or wheels of a carriage as it made contact with wires stretched across the track. He was thus the first person to photograph sequences of movement. He later created a moving image from his still sequences with his invention of the zoopraxiscope.

Muybridge took more than 20,000 photographs from 1884-85 of men, women, children, animals and birds in almost every conceivable type of movement resulting in the most comprehensive analysis of movement. His work was published under the title “Animal Locomotion” in 1887 and is still used widely today as a source of illustration and reference.

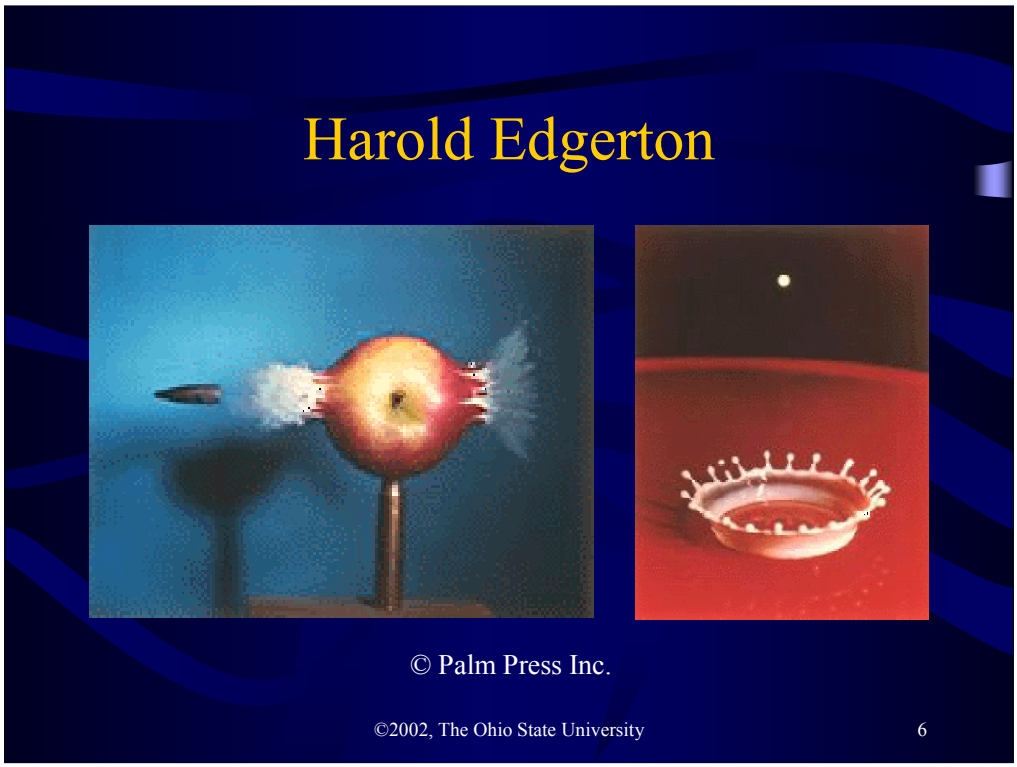
Etienne-Jules Marey

- First person to analyze human and animal motion with video
- Created photographic gun

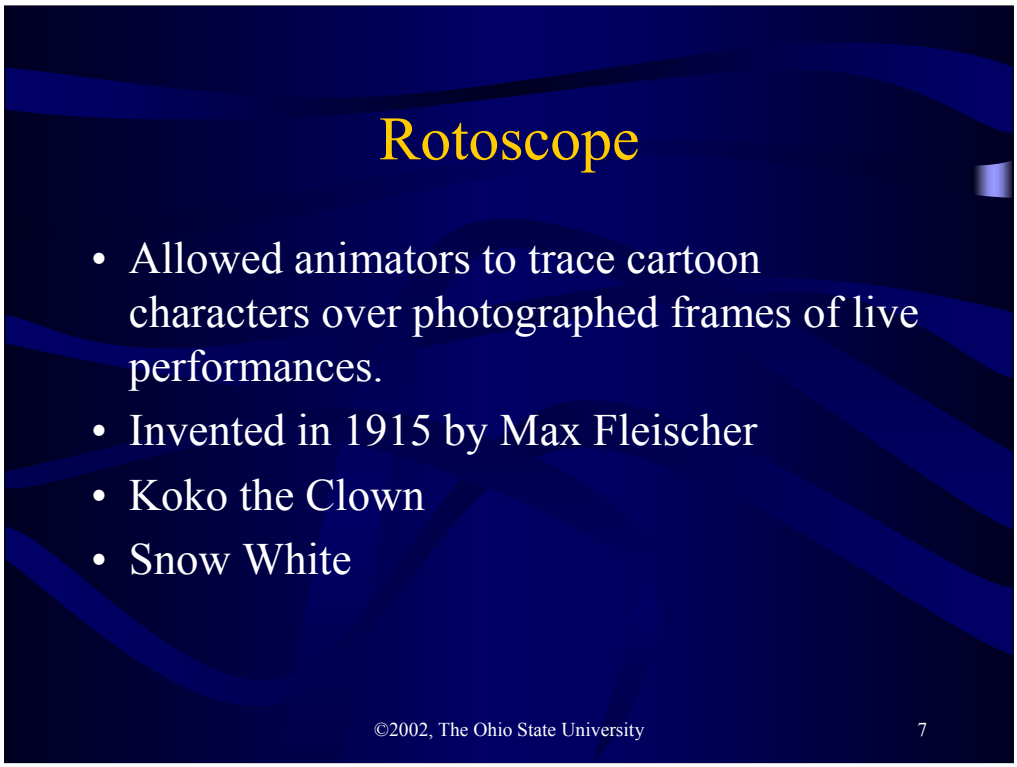


Etienne-Jules Marey was a French physician, inventor, photographer, and Professor of Natural History who specialised in human and animal physiology. He first became introduced to the study of motion when he spoke with Muybridge in 1880 about his pictures of horse locomotion. When Marey discovered that Muybridge had not had any success with photographing birds in flight, he decided to tackle the problem himself. To solve this problem, Marey invented a photographic gun. It consisted of a rotating wheel with slits. When light passed through the slit, a photographic plate was exposed. The gun had a clock-mechanism so that when the shutter was tripped it made twelve exposures of 1/72nd of a second each. Marey's study of the changes in the shape of birds' wings during flight in relation to air resistance was a major contribution to current knowledge of aerodynamics.

He also invented the choronophotograph which were multiple exposures on single glass plates through a camera of his design. He conducted many human studies with his subjects wearing a black suit with metalling threadings. The subjects walked in front of black panels and their movements were recorded by one camera, on a single metal plate.

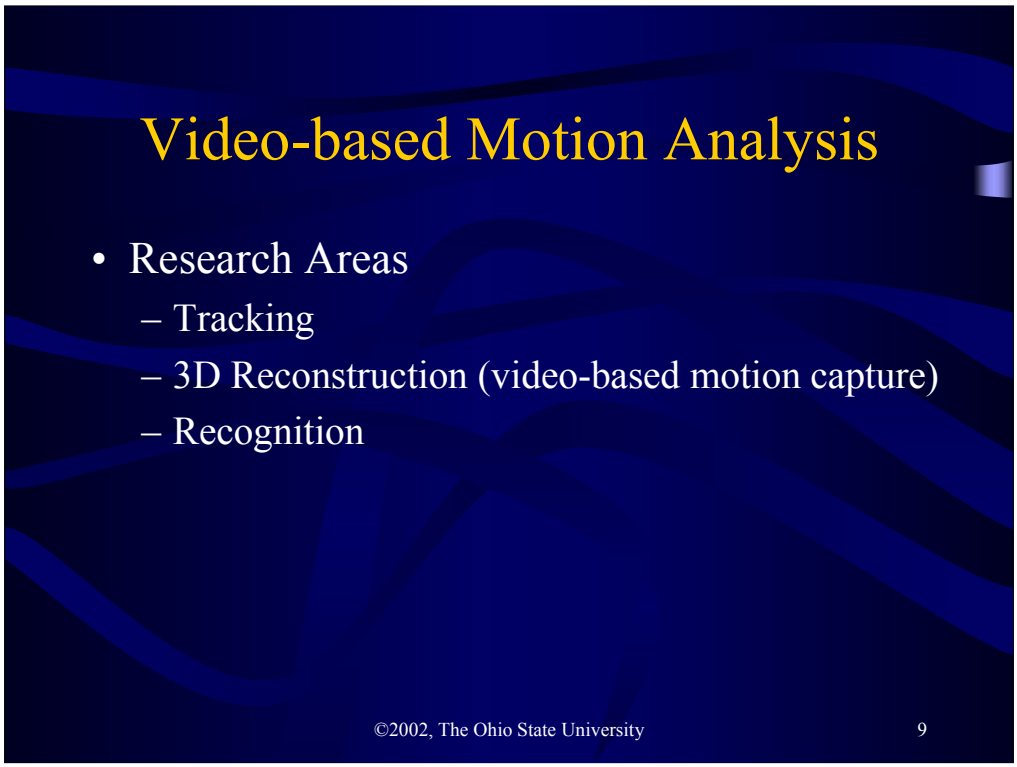


Harold Edgerton, an MIT scientist is known for developing the stroboscope and electronic flash for photographic illumination. He developed the stroboscope in 1931 for ultra-high-speed and stop motion photography. In 1932, he began taking high-speed photographs of familiar activities that move at speeds beyond the ability of the human eye to perceive. His most well known photograph being the coronet on a drop of milk. His images were revolutionary because they were taken with exposures between thousandths and up to one millionth of a second, and revealed more than the eye could see.

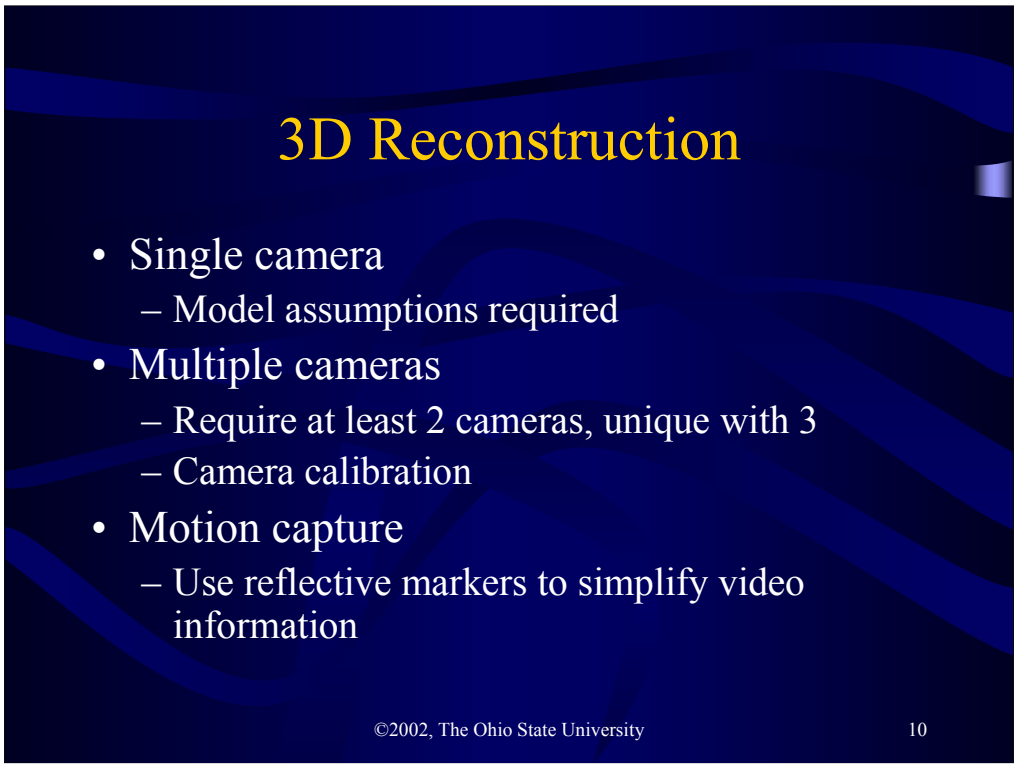


In 1915, using Muybridge's idea, cartoonist Max Fleischer created the rotoscope, a device that allowed animators to trace cartoon characters over photographed frames of live performers. A time consuming process, mainly used for human motion, it was the first time a real-life performance was used to help create an animated character. The first cartoon character to be rotoscoped was Koko the Clown. It was later used by Walt Disney, in 1937, to get realistic human motion for Snow White and the Prince. Rotoscoping is a two-dimensional approach to capture motion.

Video-based motion analysis has been a research problem in the field of computer vision for over twenty years. Research in the above three areas is hampered by several difficult problems. Various assumptions can be made to simplify the task. The first of the problems is a complex varying environment which can be reduced by requiring a static and/or uniform background, restricting the movements and number of objects and people and restricting the complexity of objects in the field of view. The second problem, segmentation, which involves extracting regions of interest from the image, is hampered by image quality, motion blur, low contrast images, and strong shadows. To make it easier, the subject may be required to wear tight fitting or colored clothes, or have high-contrast markers attached to his/her joints. The final problem, occlusion, occurs when a part of the body is obstructed by an object or other part of the body in the camera view. Early research avoided this problem by not allowing occlusion in a motion sequence but since most natural motion involves occlusion, this assumption is not made as frequently in recent research.

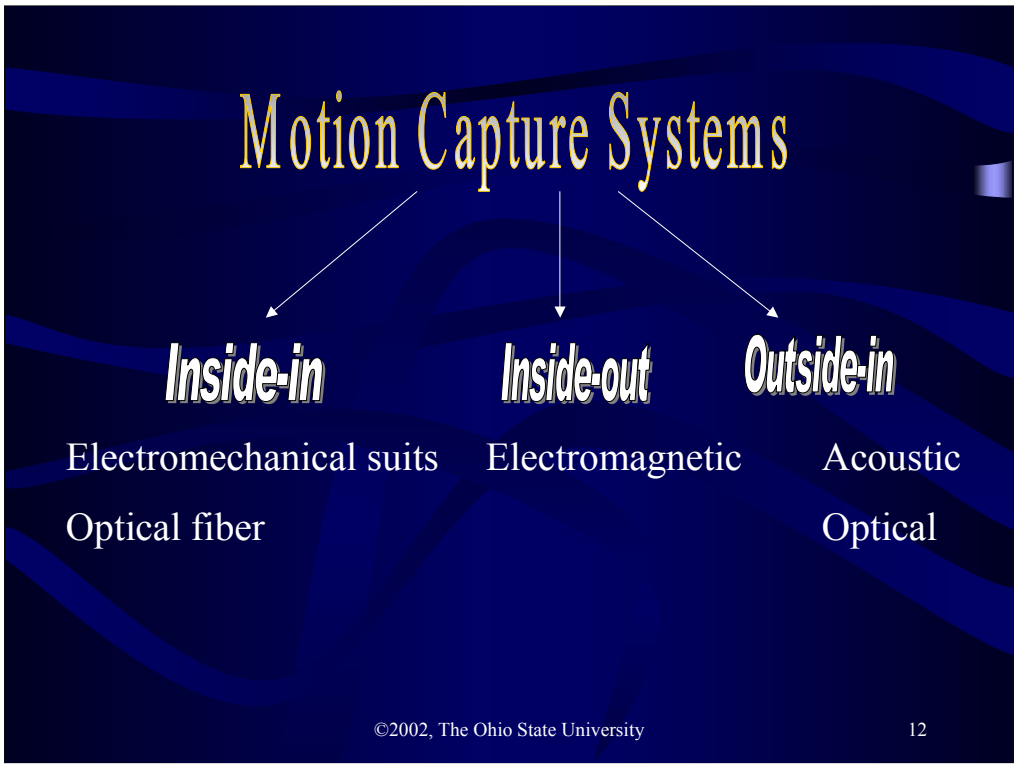


The three primary fields of research in human motion analysis are motion tracking, pose reconstruction, and motion and subject recognition. Motion tracking refers simply to the identification and location of subject characteristics from one frame to the next in the recorded human motion video and can be used in the surveillance of airports and government institutions and in monitoring traffic flow patterns. Pose reconstruction determines the actual 2D or 2D positions of the body parts in every frame of the video sequence and can be used in medical applications such as gait analysis, in the sports industry to study athletic performance, and any other motion capture applications. The final area, motion and subject recognition determines the type of motion that takes place or identifies the individual creating the motion. Recognition can further extend on either tracking or pose reconstruction using the results from either of these two steps to aid in recognizing the motion and would be useful for security applications and computer-user interfaces.



Pose reconstruction from a single camera requires model assumptions to determine 3D positions. One possibility is to use a geometric model of the human body in addition to edge and/or region features of the image sequence. The model may assist in resolving discrepancies in the reconstruction due to, for example, occlusion of body parts, or it may initiate or drive the analysis by predicting body postures in future frames. In general, the model is kept simple enough to reduce the number of parameters to be determined, but at the same time, complex enough to represent sufficient characteristics of the subject. Pose reconstruction from multiple cameras has the additional work of calibrating the different cameras and determining correspondence between the motion sequences from the different cameras. Motion capture is a specialized form of Pose Reconstruction from video. Rather than processing all the information available in a video sequence, the information is reduced by having a subject where reflective markers. It is now necessary to only locate and process the markers in a motion sequence.

Motion capture has been in use for decades for military and medical purposes. It was first used in computer graphics research in the late 1970's, early 1980's at schools such as MIT, Simon Fraser Univ., and New York Institute of Tech. Actual production use began in the mid-1980's. The use of motion capture for the chrome character known as the sexy robot in the 1985 Superbowl commercial Brilliance for the National Canned Food Information Council by Robert Abel was the first use of motion capture for 3D animation. They used a black magic marker to mark 18 joints on a dancer and then subsequently photographed multiple views as she performed the motion. In 1988, PDI created Waldo C. Graphic for the Jim Henson Hour. They used a custom eight degree of freedom input device to control position and mouth movements of a low resolution character. They were able to capture the motion in real-time in concert with real puppets. The computer image was mixed with a video feed off a camera focused on the puppets so that everyone could perform together. In post-production, Waldo was re-rendered in full resolution adding dynamic elements. Also, in 1988, deGraf/Wahrman developed Mike the Talking Head an interactive animation at Siggraph 88 for Silicon Graphics to show off the real-time capabilities of their new 4D machine. The animated character was controlled in real-time by a puppeteer during the conference. In 1989, Kleiser-Walczak used motion capture for the digital actor Dozo in the music video Don't Touch Me.



Motion capture systems can be divided into three areas based on the location of sources and sensors. Inside-in systems have the sensors and sources on the body. An example of this is an electromechanical suit where the sensors are potentiometers and the sources are the actual joints inside the body. The second area, inside-out has sensors on the body that collect external sources such as in electromagnetic systems where the sensors move in an externally generated electromagnetic field. The final area outside-in have external sensors which collect data from sources on the body such as in optical systems where the cameras are the sensors and the reflective markers are the sources.

Electromechanical suits

- Linked structures
- Potentiometers determine rotation of each link

Electromechanical suits are based on a set of linked structures which are attached to the performer's body. The structures are linked using angular measurement devices at the joints. Electromechanical suits do not have to deal with the occlusion problem. The suits are portable and are also less expensive than optical or electromagnetic systems. With multiple suits, it is possible to capture multiple performances. However, the person's movement is constrained by the armature and the sensors are fixed and cannot be changed without changing the armature.

Optical fiber system

- Fiber-optic sensor
- Rotation based on transmitted light

Optical fiber systems such as data gloves use a fiber-optic sensor along the fingers. As the fingers bend, they bend the fiber and the transmitted light is attenuated. The finger joint rotation measurements are based on the strength of the attenuated light.

Electromagnetic systems

- Electromagnetic sensors placed on joints of moving object
- Measures orientation and position of sensor with respect to electromagnetic field generated by transmitter

Electromagnetic systems use a centrally located transmitter which emits an electromagnetic field and a set of receivers which are attached to various parts of the body. The receivers can measure their spatial relationship, both their position and orientation, with respect to the transmitter. This system does not have to worry about occlusion problems and both position and orientation information is collected, but the person depending on the system may be constrained by cables and the capture volume is usually not as large as in optical systems.

Acoustic motion capture involves the use of a triad of audio receivers. Audio transmitters are strapped to various parts of the performers body. The transmitters are sequentially triggered to output a "click" and each receiver measures the time it takes for the sound to travel from each transmitter. The calculated distance of the three receivers is triangulated to provide a point in 3D space. The main problem with this approach is the sequential nature of the data positions it creates. Ideally, we would like a snap shot of a performers skeletal position. The position data thus obtained is typically applied to an inverse kinematics system which in turn drives an animated skeleton.

One of the big advantages of this method is the lack of occlusion problems normally associated with optical systems. Major disadvantages include the hindrance of the cables, the current systems do not support enough transmitters to accurately capture a detailed performance and the size of the capture area, which is limited by the speed of sound in air and the number of transmitters. In addition, the accuracy of this approach can sometimes be affected by spurious sound reflections.

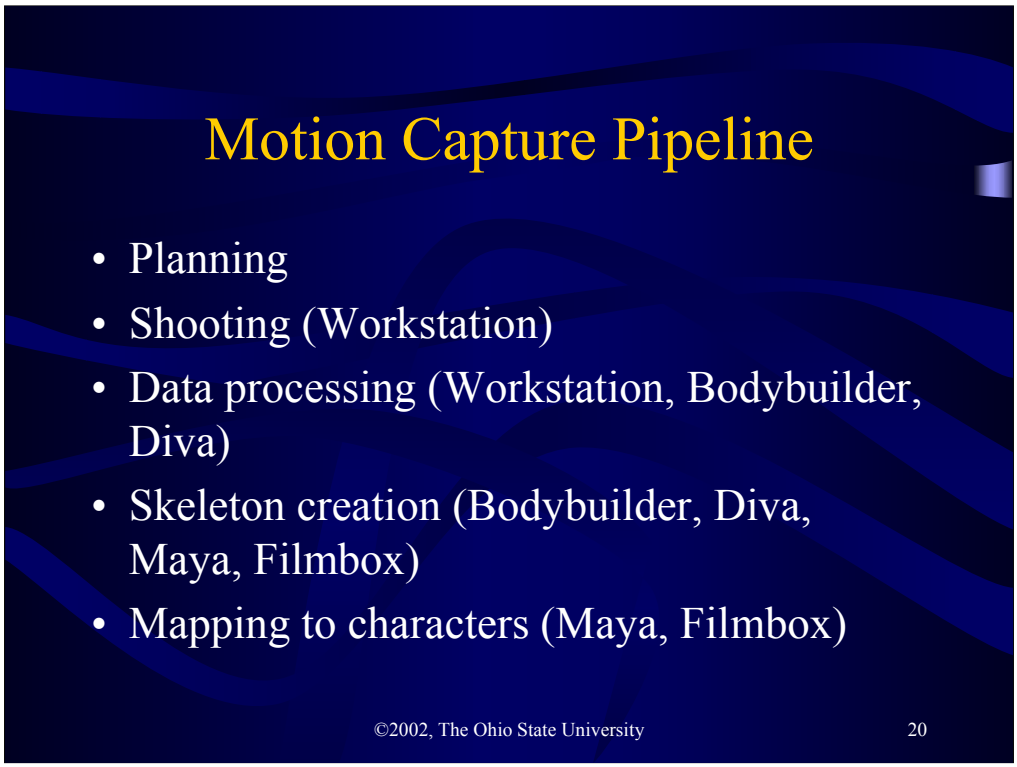
Optical motion capture system

- Reflective markers
- Multiple cameras digitize different views of performance

An optical system consists of a computer controlling several light-sensitive cameras placed strategically around the capture space. The cameras capture the light in the field of view and measure the intensity of light for each pixel in the image. The performer wears spherical markers that are covered with a highly reflective tape. The cameras have shutter synchronization and are usually fitted with their own light source that creates a directional reflection from the markers. The views from the different cameras must be calibrated so that the computer knows the location of the different cameras and can determine 3D positions of the markers. You need to have at least three cameras to determine a 3D point in space from 2D images. The advantages of an optical system include large performance areas proportional to the number of cameras, markers can be moved depending on the object to be captured, and the performer is not seriously constrained by the markers. The major disadvantages are extensive post-processing, a controlled environment, and occlusion of markers which to some extent is overcome by having redundant camera coverage from all sides.

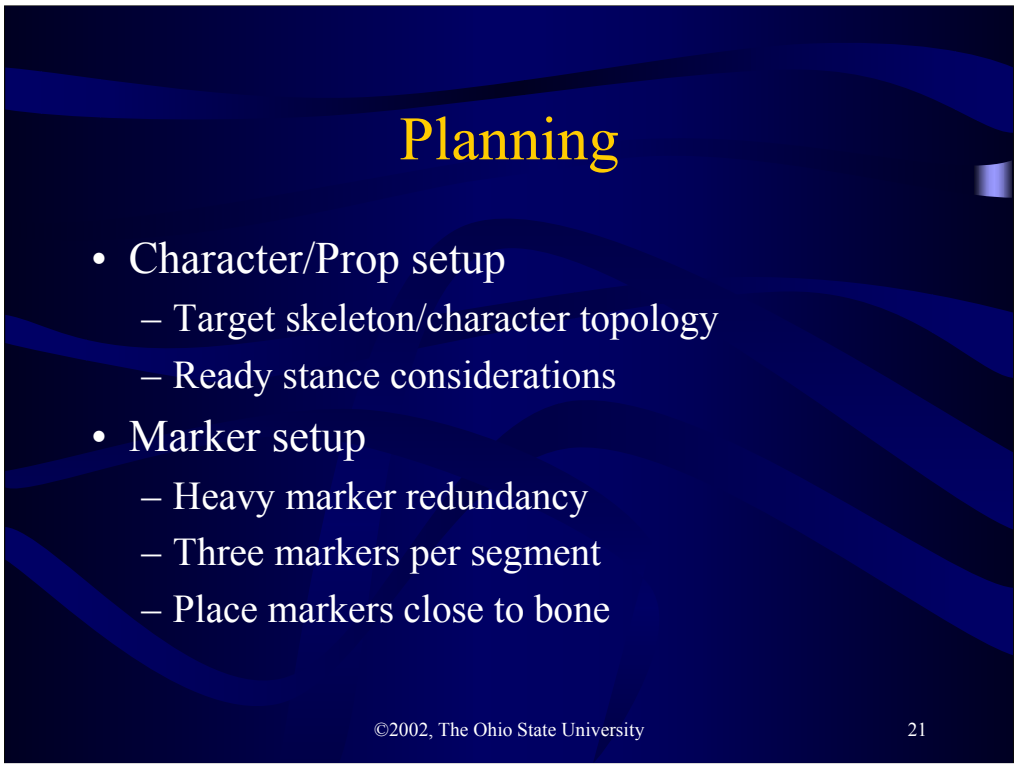
Motion capture can be used in the field of Medicine for gait analysis, sports medicine, and in prosthetic design. It can be used in sports to improve athletic performance. In Entertainment, it can be used for video games, television and feature films. Motion capture can be used to produce a reconstructive video of events in criminal law. For example, it was used to recreate events of the murder of Nicole Simpson and Ronald Goldman. The video wasn't used as actual evidence in that trial. For a reconstructive video to be admissible as evidence in a trial, the animation must be plain and supported by accompanying testimony from animation creators during trial or deposition.

Motion capture has been used extensively in feature films in the last ten years. Some examples of its use include for digital extras in the Titanic, digital stunts in Batman and Robin, facial capture in the Real Adventures of Johnny Quest, lead actor/fight scenes in The Mummy and The Mummy Returns and finally for multiple lead characters in Final Fantasy.

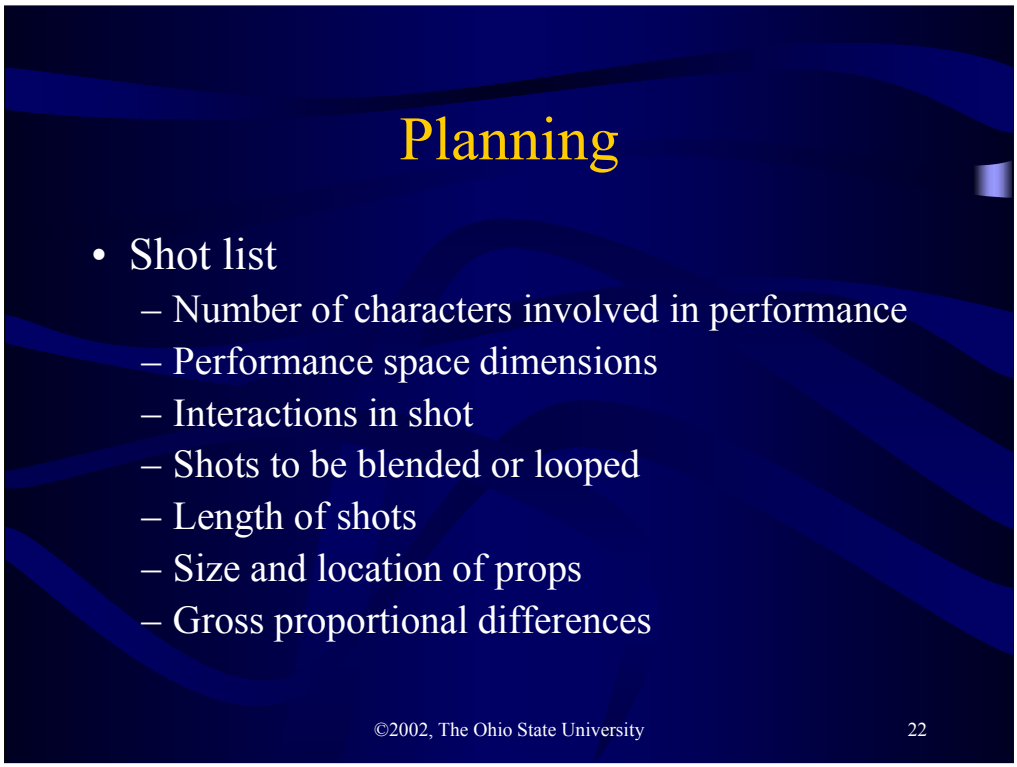


The motion capture pipeline consists of planning, shooting, data processing, skeleton creation, and mapping to characters. The most important point about using motion capture is to avoid problems by planning well ahead.

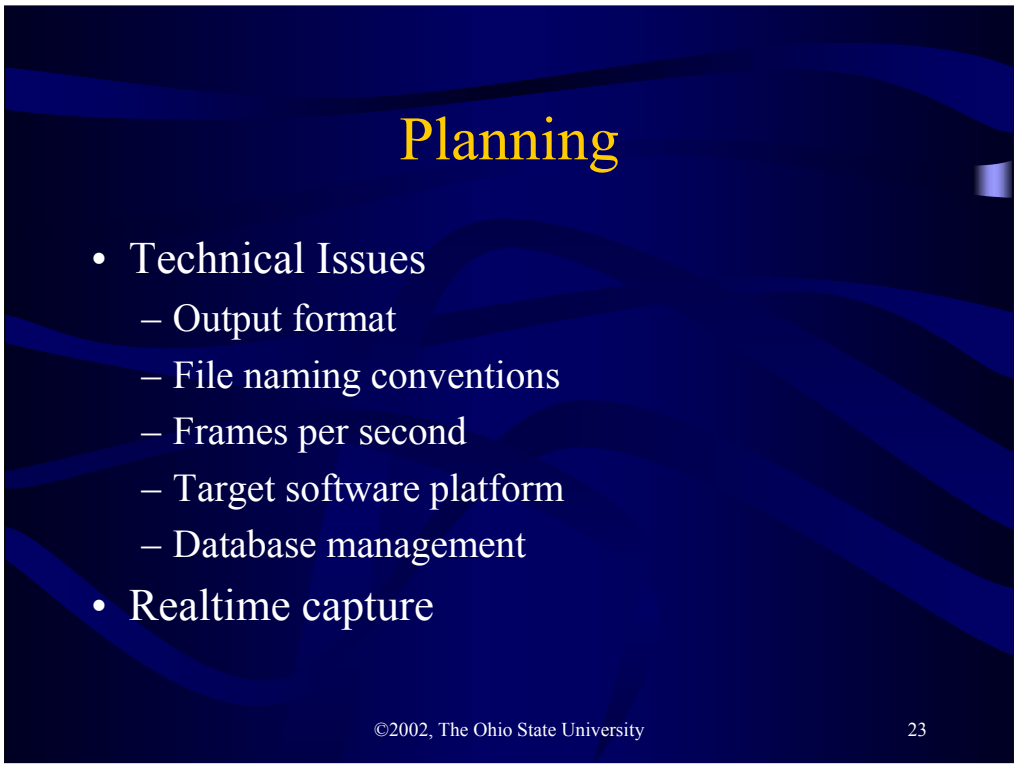
After the shoot, data processing consists of reconstructing the data from the different camera views to produce 3D positional data and labeling the markers. Once this has been done any noise in the data needs to be filtered and gaps in the data due to occlusion of markers needs to be filled.



Planning a motion capture shoot involves understanding the objectives of the shoot and how the data is to be used. You should have a storyboard or game design ready and determine the character and marker setup of any characters or props to be motion captured. Character setup deals with locations of joints or bones in the body of the character that will provide final motion and deformations. Marker setup pertains to the locations of markers that are used to collect the data. Character setup depends on marker setup because data collected must be enough to calculate all the information needed by joints. You need to have at least the character setup design decided to come up with a base pose and marker setup. When designing the marker set for the character allow for marker redundancy, at least three markers per rigid segments and make sure the markers are placed close to the bone to reduce marker sliding on the skin. Also, markers should be placed on rigid segments asymmetrically so that, for example, the software can determine the left arm from the right arm.



A shot list must be prepared in advance and should include the number of character involved in the performance and any interactions between the characters or props. The dimensions of the performance space required for the shot and the length of the shot should also be indicated. Any shots to be blended or looped and whether the blending will occur in the beginning or end of the shot must also be determined. Finally the size and location of props in the performance space and any gross proportional differences in the characters must also be listed. For example, if the character is wearing armor, the performer might not to swing his arms slightly away from his body so that the arm doesn't go through the armor on the character. Ideally, it's better to have the performer wearing a similarly weighted costume as the character will be wearing. It is visually obvious if a person is carrying a light-weight object instead of a heavy-weight object.



Some technical issues to keep in mind are the file formats that your target software is capable of handling, file naming conventions and the number of frames per second required. File naming conventions can include the character the file belongs to, the setup used, and the action implemented. It is usually best to use a capture rate that is a multiple of the frames per second required. Finally, you will be handling a lot of files and it is useful to have some kind of database management system implemented.

Project 1 Assignment

- Create an animation storyboard requiring motion capture
  - Must include at least one unusual elements such as a prop, multiple characters, non-human data, facial or hands capture
- Plan the motion capture session
  - Should answer all questions in planning worksheet

References

1. Sturman, David J., “ A Brief History of Motion Capture for Computer Character Animation”, “Character Motion Systems”, Siggraph 94, Course 9.
2. Menache, Alberto, “Understanding Motion Capture for Computer Animation and Video Games”
3. http://www.kingston.ac.uk/Muybridge/

References

4. http://www.anotherscene.com/cinema/firsts/marey.html
5. http://cmp1.ucr.edu/exhibitions/edgerton/edgerton.html
6. Siggraph 2001 Course 51: