Animation From Observation: Motion Capture and Motion Editing

Michael Gleicher
University of Wisconsin, Madison

Animation is a uniquely expressive art form: it provides the creator with control over both the appearance and the movement of characters and objects. This gives artists tremendous freedom, which when well used, can create works with tremendous impact. This freedom, however, also becomes a curse: while everything can be controlled, everything must be controlled. Control over the movement of objects is a difficult task, requiring skill and labor.

Since the earliest days of the art form [Lutz], animators have observed the movement of real creatures in order to create animated motion. Sometimes, this simply takes the form of an artist carefully observing nature for inspiration. Another process is to transfer the movement from a recording of the movement to the animated objects. The earliest mechanism for doing this was the Rotoscope, a device that projected frames of film onto the animator's workspace, providing the animator with a guide for their drawings.

Computer animation brings the potential for automating the process of creating animated motion from observations of real moving objects. Optical, mechanical, or magnetic sensors record the movements that can then be transferred to animated characters. This process is commonly referred to as motion capture, although the act of "capturing the motion" is only one aspect of creating animation from observations of real motion.

This article attempts to provide an overview of the process of creating animated motion from observations of real moving objects, and to discuss the potential for computer vision to contribute to this. My view is that the needs of the entire process create requirements on the individual steps; that motion capture for animation is most useful when the use of that data, including mapping and editing, is considered. The task of creating animation has some unique demands, and that only by considering these demands can a capture method be a useful tool for motion creation.

This article is organized as follows. We begin with a discussion of the use of motion capture to create motion for animation, and look at the alternatives. We then consider the entire process of creating animation from motion capture, and consider some of these steps in detail. Specifically, we examine the current technologies for capture and issues in working with motion data. We conclude by discussing the opportunities for computer vision in the process.

Within the animation community, there is historically a tension between animators and motion capture technicians/users [Cameron]. This tension comes from many factors, some of them real and some of them perceived. The two main sources of this tension are unrealistic expectations about what motion capture can do (that it can automatically produce motion that displaces animators), and that motion capture technology development has not considered the use of the data, leaving animators with data that is difficult to deal with.

Motion Capture vs. Animation from Observation

Motion capture is different from the process of creating animation from observations. For one, motion capture may be done for a variety of reasons besides animation, such as biomedical analysis, surveillance, sports performance analysis, or as an input mechanism for human-computer interaction. Each of these tasks has similarities and differences with the problems of creating animation. At the first stage of each, there is a need to create the observations that are then interpreted, e.g. capture the motions. Many of the methods used in animation have their roots in the bio-mechanical or medical domains.

Capturing the motion is only part of the problem of using this data to create animation. Commonly, the term motion capture is used to describe the whole process. This has the problem that it neglects other aspects of the task, and sets up some unreasonable expectations about how much work needs to be done to move from the sensor data to animation.

Let's begin with the question of what is capture anyway. In a sense, pointing a video camera at a person captures their motion. We can play it back and see what they did. For some reason, this is not what we commonly mean by motion capture. The distinction (for me at least) is that motion capture creates a representation that distills the motion from the appearance; that it encodes the motion in a form that is suitable for the kinds of processing or analysis that we

need to perform. This definition of motion capture is dependent on what we are going to do with the result.

Motion capture for animation implies that we will somehow be changing something about what we have recorded–if we did not intend to change something, we could have simply replayed a video. Almost always, we will at least change the character to which the motion is applied from a real person to some graphical model. By definition, to animate means to bring to life, so technically, it is the act of making a lifeless object (a graphics model) move that makes what we're doing animation.

There is a range of types of motion capture for animation. One distinction is between real-time, on-line systems where the animation is produced instantly, and systems that are not real time. While the former category is best known in applications where it is required, such as creating characters for live broadcasts or interactive exhibits, it is also often useful for creating traditional animation as well. Even if the final result will require adjustment and production, instant feedback to the performer is useful. The production of real-time animation from captured motion is sometimes referred to as performance animation or digital puppetteering.

Another distinction in motion capture is between capturing facial motion and capturing body motion. Our focus in this article is on full-body motion. Facial motion capture has a similar set of issues with a slightly different set of challenges than body animation.

Motion Capture vs. Animation

On-line motion capture is unique in that it is an application for which there is no alternative. For off-line production, however, motion capture is only one of several ways to create motion for animation. Understanding the alternatives is useful to see where motion capture is most useful, and what it must be able to do to serve as a mechanism for creating animated motion. Taxonomies of motion creation, including [Hodgins], usually divide methods into three categories: manual specification, procedural and simulation, and motion capture.

Traditionally, motion for animation has been created by specifying the position of objects at each instant in time [Lutz]. These methods became highly evolved as the art developed [TJ]. Manual specification has the obvious drawback of being laborious, but also requires a great deal of skill to create convincing motion by specifying a series of individual poses as properties of the motion are created over many individual poses. While computers can reduce some of the labor by

automatically interpolating between keyframes, manual specification of motion still requires talent and training [Lasseter]. It is particularly difficult to create motions that are realistic and/or accurately mimic subtle characteristics, such as a particular person.

Another strategy uses algorithmic or simulation methods to generate motions based on descriptions of goals. While such methods have the promise of generating motions for non-experts by allowing them to simply specify their needs, they are, at present, of limited use, as there has been no systematic way provided to create new behaviors. One key problem facing algorithmic methods is how to describe a complicated motion or a subtle nuance.

An alternative to the above three methods is not a motion creation method per se, but rather is to avoid creating a new motion. Instead the needed motion can be created by re-using an existing motion. In practice, such an approach requires two pieces: a library of motions to re-use, and techniques to adapt motions to new needs. The limitations of this approach come from its two components, the library of motions available to adapt, and the quality of the tools available for adapting motions.

In a performance setting, there really is no alternative for motion capture. For off-line production, motion capture must provide an advantage over other available methods. In order to be a viable alternative, motion capture must provide a sufficient quality of service, both in terms of quality of resulting motions and in range. For example, if motion capture does not provide sufficient fidelity to distinguish the subtle differences between different performers, a standard motion from a database may be sufficient. Or, if a motion capture system can only capture a limited range of motions, this range may be covered by a library. The existing approaches to motion creation set a high standard that a new tool must meet.

Motion Capture for Animation

The steps in creating animation from observation are:

1. Plan the motion capture shoot and subsequent production. Good planning is amazingly important to make motion capture work in practice [Kines].
2. Capture the motion.
3. Clean the data.
4. Edit the motions.
5. Map the motions to the animated characters.

The order of steps 4 and 5 are often varied, depending on the tools. Sometimes, these steps are actually iterated.

While the production pipeline provides opportunities to fix problems created in earlier stages, it also means that these problems cause additional work later on. Therefore, we prefer motion capture to have problems that are easily addressed in later stages than to have fewer, but harder to correct problems.

Capturing the Motion

A variety of methods have been used successfully to "capture" motions. At one level, the actual technology for sensing and recording a person's performance is irrelevant as different methods should lead to similar results. However, each different approach has a different set of tradeoffs and exhibits a different set of issues in its results. While the vendors of various capture systems are continually improving all of the varieties of systems, the experience of users in practice is still dominated by the limitations of particular devices.

Motion capture for animation has a long, fascinating, but poorly documented history [WC]. The earliest motion capture systems used mechanical armatures to measure joint angles. Early examples used goniometric harnesses designed for medical analysis to drive analog computers. More recently, the use of mechanical technology has primarily been used to create puppets. Systems designed to track a human figures motion require a mechanical skeleton to be strapped onto the performer. Modern implementations of this approach use clever mechanisms to reduce the encumbrance.

Magnetic motion capture technologies use transmitters that establish magnetic fields within a space, and then use sensors that can determine their position and orientation within the space based on these fields. Early versions of magnetic systems were plagued by practical problems: the sensors required cables that encumbered performers, there was significant sensor error creating noise and drift, the fields were of limited range, and the magnetic fields were easily interfered with by metal objects in the space. Modern magnetic capture systems address these issues to a degree: wireless versions place radio transmitters on the performer's body and updated sensors provide better performance, range and robustness.

Optical tracking systems use special visual markers on the performer and a number of special cameras to determine the 3D location of the markers. Traditionally, the markers are passive objects, such as retroreflective spheres, and the cameras are high-speed, monochrome devices tuned to sense a specific color of light. Optical systems require multiple cameras to see a marker in order to triangulate its position, and may "drop" markers due to occlusion.

State of the art optical systems often use many cameras (sometimes as many as 24) in an effort to minimize the risk of markers not being seen by enough cameras.

One challenge of an optical system is that while they are able to see where the markers are, they have no method to know which marker is which. Unlike a magnetic system, where each marker has its own sensor data channel, an optical system must determine the correspondence of markers between frames. Typically, this is done in post-processing software based on continuity of positions. Optical systems typically prefer high frame rates to create this continuity, even if the resulting data will be down-sampled. While analysis techniques are improving, software techniques are still imperfect and require manual cleanup. Hardware solutions use active markers such as miniature LEDs to disambiguate markers.

Because optical capture systems must address lost markers due to occlusion and correspondence as post-processes, magnetic systems have traditionally been the preference for performance animation. Improved software for optical processing is changing this. Similarly, both technologies are evolving rapidly, changing many of the historical tradeoffs in their relative price-performance.

We make a somewhat arbitrary distinction between optical capture technology and computer vision-based capture technologies. We define a vision-based technology as one that can analyze "standard" video streams, performing some form of image analysis to determine what the performer is doing. Optical technologies usually provide engineering solutions to standard vision issues such as tracking and identification, for examples using special cameras and lighting that make markers obvious. The first commercial, video-based, full-body motion capture system is available from Peak Performance [Peak]. To date, however, vision-based capture has not been a successful tool for animation. We will discuss this in a later section.

Motion Editing and Motion Capture

Motion capture techniques (ideally) should provide wonderful motions - why should there be a need to change them? If everything was working correctly, motion capture data should be an accurate reflection of the reality of a desired performance. Yet the discussion of how to change motions once we have them always seems to be a big part of the use of motion capture.

A common misconception is that the importance of motion editing for motion capture comes from the fact that motion capture is imperfect, and that tools are

needed to clean-up the motion after the fact. Even when the motion capture data perfectly represents a desired performance, there is often a need to make alterations to the motion, for reasons including:

- **Re-use**: motion capture data exactly records an event. If we want to re-use the data for something slightly different, say a different character or a different action, we need to edit the data.
- **Creating Infeasible Motions**: because motion capture records real events, some editing is required to make "impossible" actions happen.
- **Imperfections of reality**: real motion isn't perfect. Performers don't exactly hit their marks and repetitive motions are not exactly cyclic.
- **Change of intent**: unfortunately, we can't always predict what motion we will need, and even if we do, someone might change their mind as to what is desired after the fact.
- **Addition of "secondary" motion**: such as the movement of soft tissue or clothing.

Motion editing problems and techniques are not unique to motion capture and can be applied to motion created with other methods, such as key-framing and simulation, as well.

There are a number of issues that make working with motion capture data more difficult than working with traditionally animated motion. These issues transcend the technology used to capture the motion.

- The data is most certainly inconvenient for editing. Motion capture systems typically provide a pose for every sample or frame of the motion, not just at important instants in time. This means that a lot of data must be changed to make an edit.
- There is nothing but the data to describe the properties of the motion. There is little indication in the data to show what the important properties of the motion are, and what should be changed to effect the motion, nor is there an animator familiar with the "why" of the motion.
- Sensor errors and other failures lead to "dirt" in motion, requiring clean-up. What makes this a challenging problem is that because we don't have an exact record of what happened (if so, we could use that instead of our dirty data), it is difficult to know when our data is wrong, and even more difficult to know what to replace it with.

Computer Vision and Motion Capture

There is a growing segment of the computer vision community that is interested in the problem of analyzing images of people in motion. The applications are varied, e.g., surveillance, input for user interfaces, and bio-mechanical analysis. Just as traditional motion capture techniques have been applied from these other domains into animation, video analysis offers an attractive device for the creation of animated motion.

The potential for vision-based motion capture is great: conventional video technology is more accessible, less costly, less encumbering of the performers, and works in a wider variety of environments than current capture technologies. If standard video can be analyzed, legacy footage can be processed to create animation. However, these possibilities require a capture technology that can provide the fidelity and quality that animation applications require. These demands are different than the applications on which vision researchers have focused.

Tracking human motion has been an important topic in computer vision. However, for most applications creating a 3D reconstruction of the motion is not required. For example, [YB] create 2D representations of motion for action identification, and a variety of interactive demonstrations have been based on silhouette tracking [DB]. Applications, such as user interfaces, require real-time performance, even at the expense of fidelity *[Freeman, this issue]*.

Several researchers have focussed on the problem of determining articulated figure configurations from images. The earliest attempts [OB] reconstructed a graphical model of the figure and used this in a feedback loop. Similar approaches have been demonstrated by [GD] and [KM]. More recently, [BM] and [YSK] have described systems that compute the motion of articulated figures using differential optical flow techniques. To date, none of these efforts have provided motion of the fidelity demanded for animation production, and they have only demonstrated short simple motions (with the exception of [GD] which required a very controlled situation).

Efforts to perform facial motion capture using vision technology are much more established. Because of the limited range of motion and slower movement rates, video has been a much more workable methodology for facial capture. Williams' early facial capture [Williams] was performed using a single video camera and a mirror to generate the multiple views required for stereo reconstruction. Terzopolous and Waters [TW] tracked marks on a performer's face using snakes, and used these curves to drive a muscle-based facial model. In fact, with recent advances in facial tracking, it seems that the largest challenge in creating vision-based motion capture is to provide data in a form usable by facial animators.

Another exciting opportunity where computer vision and motion capture may link is the use of vision technologies to assist in the process of using animation data. The experience of computer vision research in dealing with noisy unidentified signals can apply to motion data as well as images. Some early examples include the use of robust statistics for noise reduction [BRRP], and the automatic identification of constraints [BB].

As we mentioned earlier, the demands of computer animation, and alternative techniques for motion creation, place a very high bar for motion capture techniques. To date, vision techniques have fallen far short.

References

[BB] R. Bindiganavale and N. Badler, ''Motion abstraction and mapping with spatial constraints," Workshop on Motion Capture Technology, Geneva, Switzerland, November 1998.

[BM] C. Bregler and J. Malik. Tracking People with Twists and Exponential Maps.,Proceedings of Computer Vision and Pattern Recognition Conference, 1998.

[BRRP] B. Bodenheimer, C. Rose, S. Rosenthal, and J. Pella. The Process of Motion Capture: Dealing with the Data *Computer Animation and Simulation '97. Proceedings of the Eurographics Workshop.*

[Cameron] Gordon Cameron (organizer). Motion Capture and CG Character Animation. SIGGRAPH '97 Panel.

[DB] James Davis and Aaron Bobick. A Robust Human-Silhouette Extraction Technique for Interactive Virtual Environments. *IFIP Workshop on Modeling and Motion Capture Techniques for Virtual Environments, 1998.)*

[GD] D. Gavrila and L. Davis. 3-D model-based tracking of humans in action: a multi-view approach. Proceedings of Computer Vision and Pattern Recognition Conference, 1997.

[Hodgins] Jessica Hodgins. Animating Human Motion, *Scientific American*, March, 1998.

[KM] Ioannis A. Kakadiaris and Dimitri Metaxas. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection, Proceedings of Computer Vision and Pattern Recognition Conference, pp. 81-87, 1996.

[Kines]. Melianthe Kines. Planning a Motion Capture Shoot. *Game Developer*, September, 1998.

[Lasseter] John Lasseter. Principles of Traditional Animation applied to 3D computer animation. SIGGRAPH '87.

[Lutz] E. G. Lutz. Animated Cartonns: *How They are Made, Their Origin, and Their Development.* Charles Scribner's Sons, 1920. Reprinted by Applewood Books, 1998.

[OB] J. O'Rourke and N. Badler, Model-based image analysis of human motion using constraint propagation, IEEE Trans. on Pattern Analysis and Machine Intelligence 2(6), Nov. 1980, pp. 522-536

[Peak] Peak Performance Technologies, Inc. http://www.peakperform.com/

[TJ] Frank Thomas and Ollie Johnson. *Disney Animation: The Illusion of Life*. Abbeville Press, 1981.

[TW] D. Terzopoulos, K. Waters, Analysis and synthesis of facial image sequences using physical and anatomical models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(6), June 1993, pp. 569-579.

[Williams] Lance Williams. Performance-Driven Facial Animation. SIGGRAPH '90.

[WC] Lance Williams and Frank Crow. Personal Communication, 1999.

[YB] Yaser Yacoob and Michael Black. Parameterized Modeling and Recognition of Activities, Journal of Computer Vision and Image Understanding, 73(2), 1999, 232-247.

[YSK] M. Yamamoto, A. Sato, and S. Kawada. Incremental Tracking of Human Actions from Multiple Views. Proceedings Proceedings of Computer Vision and Pattern Recognition Conference, 1998.