Chris Bregler

Motion Capture Technology for Entertainment

The year 2007 started with two big bangs for LucasFilm's Industrial Light and Magic (ILM), the motion picture visual effects company started by George Lucas in 1975, in two eagerly awaited Hollywood events. In January, ILM won the "Scientific and Engineering Award" of the Academy of Motion Picture Arts and Sciences for their image-based modeling work applied to visual effects in the film industry. In February, ILM took home the Oscar for "Best Visual Effects" for *Pirates of the Caribbean: Dead Man's Chest*, a milestone achievement for ILM after their last Oscar for *Forrest Gump* in 1994 and many previous Oscars in their 32-year history of pioneering computer-based visual effects. In this column, we comment on the significance of these awards from a technical standpoint and the goal, challenges, and status of motion capture technology.

THE SIGNIFICANCE

The ILM awards are in line with the advancement of the entire visual effects industry and, in particular, the progress towards one specific goal—the creation of artificial humans and other creatures that are so realistic one can say that they passed the Turing test of computer graphics. That is, the creation of computer graphics generated humans that cannot be distinguished from film recordings. As old as the invention of film itself, this goal has always remained tantalizingly out of reach.

While special effects designers have been chasing this fantasy for 30 years, this might be the decade that we reach this momentous goal. One of the secret ingredients that will get us there may be motion capture (MoCap), defined as a technology that allows us to record human motion with sensors and to digitally map the motion to computer-generated creatures. MoCap has been both hyped and criticized since people started experimenting with computerized motion-recording devices in the 1970s. Whether reviled by animation purists as a shortcut or exalted as the solution to traditional film and animation shortcomings, MoCap's controversial applications are on the brink of transforming contemporary entertainment.

THE ILM AWARDS ARE IN LINE WITH ONE SPECIFIC GOAL—THE CREATION OF ARTIFICIAL HUMANS AND OTHER CREATURES THAT ARE SO REALISTIC ONE CAN SAY THAT THEY PASSED THE TURING TEST OF COMPUTER GRAPHICS.

THE CONTROVERSY

The controversy around MoCap can be traced back to related techniques in the 1920s, so-called "rotoscoping." This technique entailed projecting live-action footage onto cell animation drawing tables and then tracing the motion onto cartoon characters. Walt Disney never publicly admitted to this, but when the Disney archives were recently made public, most of Snow White's and all of the Prince's motions were shown to be traced from actors' motions in film recordings. This was considered "cheating" in the animation world, where it was thought that everything should be conjured up by the imagination. "Good animation" was considered to be a rare and hard-to-acquire art form that followed the famous drawing and animation principles of cartoon physics invented by Disney. Rotoscoping was thought to be "cheap animation" that lacked expressiveness. If one of animation's main principles is exaggeration—everything needs to be bigger than life, not just a copy of life—rotoscoping, as precursor of MoCap, was considered even less than life because many important subtleties got lost in the process.

As in the 1920s, in the late 1980s and early 1990s when early MoCap (or semiautomatic rotoscoping) effects appeared in feature films like *Total Recall*, the century-long controversy was revived on different battlegrounds in the visual effects and computer animation industry. At conferences like SIGGRAPH, there were heated discussions between the lead animators from Pixar, Disney, Dreamworks, and film and animation schools (who argued for the necessity of creating engaging and expressive characters) and the effects experts from ILM, Digital Domain, Sony, and others (who faced the challenge of creating photo-realistic stunts and argued that animators can never animate real people realistically, and especially not in the fast-paced time frame of their production schedules). So what exactly is this technology that has divided the industry so powerfully for so many years?

(continued on page 156)

Digital Object Identifier 10.1109/MSP.2007.906023

THE TECHNOLOGY

Digitally capturing the movement of humans can be done in many ways. Most techniques can be classified as either marker-based or marker-free capture. In marker-based capture, reflections on markers that are placed on a body suit are recorded and these points are automatically visualized. In marker-free capture, thousands of natural points are captured, as in (for instance), a real-time three-dimensional (3-D) scan.

Both of these methods have their pros and cons. Most MoCap systems that have been used over the past decade are marker based. In such systems, usually between 6 and 50 cameras are placed on the walls and ceilings of a recording studio. Infrared or visible light sources are mounted next to each camera. The projected light is reflected with maximum energy by the small "retro-reflective" markers back into the cameras, and therefore can be easily detected by further processing of the camera output. Triangulating the tracked two-dimensional (2-D) marker locations of each camera allows for very accurate 3-D reconstruction of the markers in the recording studio.

For body segment (such as upper and lower arms and legs) MoCap, each body segment has a few such markers (usually a total of 50 markers are used for the entire body) and, in the next pipeline step, a digital human skeleton can be fitted to the marker motions.

For human face capture, a number of tinier markers are placed at important facial features—for instance, around the mouth, cheeks, and eyebrows. Although the marker tracking itself is very accurate and fast, very often the recorded motions look a bit robotic because there are many subtle deformations between the markers (at the spine, hips, muscle bulges, and so on) that can not be captured by this technique. However, since it is perceptually very important to convey features like weight and force of the performer, the number of markers could be increased to overcome this problem. The most extreme version of this solution leads to "marker-free" capture, in which every recorded pixel is a marker, and the performers do not need to wear special suits. This has been the subject of many computer vision research projects in the academic community and new R&D projects in industry. In theory, this should lead to much better motions, but unfortunately, it is much more difficult to track pixels with arbitrary surface texture than with retro-reflective balls.

THE CONTROVERSY AROUND MOCAP CAN BE TRACED BACK TO RELATED TECHNIQUES IN THE 1920S, SO-CALLED "ROTOSCOPING."

SYSTEMS AND APPLICATIONS

The leading marker-based commercial systems have been developed by Vicon Motion Systems and Motion Analysis Corp. More recently, PhaseSpace caught up with the big players in terms of speed and accuracy, providing an alternative active marker system based on light-emitting diodes (LEDs) attached to the body. All commercial systems run now in real time and export the data to many standard computer graphics packages, such as Maya or MotionBuilder. Ten years ago, the use of MoCap was limited to experts in studios and some research universities, but nowadays, there are many large MoCap databases available for free or purchase, and smaller studios and schools can afford multicamera systems for production, teaching, and even low-budget art projects.

Big visual effects studios like ILM, Digital Domain, Weta Digital, and Sony Imageworks used MoCap in feature films such as *Titanic, Lord of the Rings,* and *Star Wars*, either for animating background motions (such as crowd scenes or wide-angle shots on *Titanic*) or for providing motions for nonhuman creatures and touching them up with the hands of traditionally trained animators. The most recent advances can be seen in *Pirates of the Caribbean: Dead Man's Chest*, for which ILM employed a new system that does not need any retroreflective markers, or in *Monster House*, where all the characters were shot using Sony's new MoCap system. A recent ILM innovation allows the film director to see a real-time previsualization of the MoCap effects right on the stage.

In addition to movie production, one of the biggest users of MoCap is the gaming industry. For instance, Electronic Arts has several in-house MoCap studios, including one of the world's largest studios in Vancouver, Canada, which is booked solid for all of Electronic Arts sports titles, like FIFA Worldcup (their Soccer game), Madden NFL (their Football game), or NBA Live (their Basketball game). Sometimes entire soccer or football teams are suited up in MoCap attire, and several stars have performed their signature moves for electronic recordings. These projects have formed a large database of motion snippets that can be re-ordered and concatenated in real time on game consoles to enact any motion during a game.

TOWARD MARKERLESS METHODS

Most of the motions recorded with these marker-based systems still have a slight feeling of diminished expressiveness. The characters usually seem to have the wrong weight or body force, and they move in a more robotic fashion than a living, breathing human. This was acceptable in *Total Recall* or *Star Wars—The Phantom Menace*, when the target characters were robots, or in *Lord of the Rings*, where animators had the time to touch up the motions in postproduction. But in features like Square-Soft's *Final Fantasy* or *Sony's Polar Express*, for which there was no time to refine the motions by hand, the characters were stilted and somewhat zombie-like. This again gave added fuel to the century-old debate between the animators and the MoCap experts.

More recent applications of MoCap suffer less from this phenomenon due to higher-resolution capture techniques and more sophisticated motion processing algorithms. For instance, Sony's *Monster House* and Weta's *King Kong*'s facial animation uses a facial action coding system (FACS), based on a method pioneered by psychologist Paul Ekman. Instead of

FIGURE 1 A new high-speed optical motion capture system [4], developed at Mitsubishi Electronic Research Laboratories (MERL), can measure 3-D motion, orientation, and incident illumination at tagged points in a scene. (Photo courtesy of Ramesh Raskar of MERL.)

directly mapping marker to marker, which creates the look of a human in a gorilla mask, motions are mapped onto the perceptual equivalent for a gorilla, resulting in more ape-like behavior [1].

Even Disney's "Secret Lab" experimented with facial capture, completely ignoring markers and estimating optical flow (pixel movement) directly from the skin, lip, and eye surfaces, and automatically animating new characters [2]. ESC Entertainment, the main visual effects house behind the *Matrix* movies, developed another markerless optical-flow-based system to capture Agent Smith's and Keanu Reeves's facial motions. Most viewers didn't notice that some faces were artificial in several stunt shots. This system is called Universal Capture and it was developed by George Borshukov (who received an Academy Award in 2000 for the famous "bullet time" scene in the first *Matrix*). He moved to Electronic Arts and applied this technique to a new high-resolution Tiger Woods game. Similar trends are occurring at ILM and LucasArts (LucasFilm's gaming unit), where techniques that were originally developed for film are now applied to games.

Several other companies came out with new markerless capture techniques, including Mova Contour, which requires subjects to paint their faces and clothes with Halloween-store make-up that shows up under black light, and Image Metrics, which does not need any paint or other instrumentation, but instead applies model-based computer vision techniques to automatically map human faces to game and film characters. Similar systems for full-body markerless capture by Organic Motion, Stanford's Biomechanics department (based on silhouette tracking and laser-scans [3]), and by MaMoCa (which uses structured light, i.e., projecting infrared patterns onto the body parts) are coming to the market.

THE NEXT FRONTIER

Despite the advances toward cheaper, faster, less-obtrusive systems, one still needs to spend at least US$80,000 to afford such a system, in addition to a dedicated special recording studio for the project and careful planning of the shoot. This year, several low-cost MoCap systems hit the mass market: Nintendo's Wii puts an infrared capture system into a kid's hand, and Sony's new Playstation 3 will soon offer a new high-speed camera for consumer prices. Others are able to function even outdoors [4], [5] as shown in Figure 1(a)–(c). After moving from the movie domain to gaming, MoCap is now broadening the very concept of what a game can be, moving it onto another scale entirely. In 1992 and 2004, the SIGGRAPH conference premiered two mass-audience games, Cinematrix and Squidball, in which thousands of players interacted and were tracked by vision and MoCap systems. Bringing this technology out of the studio and beyond the game console, these and other projects suggest that MoCap may go places in the future that are beyond even Walt Disney's imagination.

AUTHOR

Chris Bregler (bregler@courant.nyu.edu) is an associate professor at New York University. His research interests are in motion capture, computer vision, computer graphics, and pattern recognition.

REFERENCES

[1] P. Havaldar, "Performance-driven facial animation," in *Proc. SIGGRAPH 2006* (course notes), pp. 23–42 [Online]. Available: http://old.siggraph.org/publications/2006cn/course30.pdf

[2] W. Hyneman, H. Itokazu, L.Williams, and X. Zhao, "Human face project," in *Proc. SIGGRAPH 2005*, course notes [Online]. Available: portal.acm.org/ft_gateway.cfm?id=1198585&type=pdf

[3] S. Corazza, L. Mündermann, A.M. Chaudhari, T. Demattio, C. Cobelli, and T.P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach," *Annals Biomed. Eng.*, vol. 34, no. 6, pp. 1019–1029, 2006.

[4] R. Raskar, H. Nii, B. de Decker, Y. Hashimoto, J. Summet, D. Moore, Y. Zhao, J. Westhues, P. Dietz, M. Inami, S. Nayar, J. Barnwell, M. Noland, P. Bekaert, V. Branzoi, and E. Bruns, "PRAKASH: Lighting-aware motion capture using photosensing markers and multiplexed illuminators," in *Proc. SIGGRAPH 2007, ACM*, San Diego, CA, 2007.

[5] D. Vlasic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popović, "Practical motion capture in everyday surroundings," in *Proc. SIGGRAPH 2007, ACM*, San Diego, CA, 2007.

[SP]