Motion Capture for the Masses

JOSHUA A. STAPLES[1]   TIMOTHY A. DAVIS[2]

Department of Computer Science
Clemson University, USA
[1]joshstaples@gmail.com
[2]tadavis@cs.clemson.edu

Abstract: In this paper, we introduce a flexible and inexpensive means for acquiring motion capture data from a human actor or other object. The system described requires just a few pieces of hardware, while leveraging the capabilities of off-the-shelf software to track optical markers in raw film footage. The resulting data can then be applied to any virtual figure, as with motion capture data acquired by more traditional means.

Keywords: MOCAP, portable, cost-effective, motion capture, body-suit

(Received January 26, 2006 / Accepted March 7, 2006)

1. Introduction

Computerized motion capture is a process used in many industries, including biomechanics, robotics, military operations, sports, medicine, and entertainment. Software and hardware designers have created a variety of products and solutions to conquer the challenges within each industry, but these products tend to be tailored to specific purposes and are often expensive and not easily portable. The Vicon system [14], which is popular in the entertainment industry, is one example. The Vicon System provides a reliable motion capture solution, but at a fairly high cost due to the large amount of proprietary hardware and software that must be professionally installed and maintained, making it difficult for independent users with small budgets to take advantage of its features. The Vicon's physical installation would also limit any possibility of portability. Our goal, therefore, is finding a portable and affordable solution to motion capture that is appropriate and accessible to smaller operations.

To achieve this goal, we must start with the basics. A successful motion capture system typically includes several stages: camera calibration, marker selection and placement, and tracking (i.e., post-processing to convert filmed markers to 3D coordinates). The first stage, camera calibration, is extremely important since it results in the determination of several mathematical properties that represent the intrinsic and extrinsic properties of the camera, both of which are necessary for accurate tracking [2] [12]. The second stage involves selecting the correct type of markers and placing them on the actor so that they are detectable, but not so intrusive as to inhibit an actor's movements [8]. For the third stage, we film the desired action with a multi-camera set-up to capture marker locations from several perspectives, and process the data to extract these locations such that they can be applied to a virtual character [1] [15].

Of course, we would like to create a system that is able to implement the above stages in a simple, cost-effective manner. To achieve this goal, we have developed an inexpensive marker suit, along with portable camera setup and calibration pattern, that can be used in conjunction with a set of third-party software tools to track markers in an effective way. Our overarching goals have been to develop a means of motion capture that is accurate, portable, and accessible to a broad range of users.

Section 2 discusses related work in the field to provide a context in which the current work can be placed. Section 3 describes our process to acquire motion capture data in a portable and inexpensive manner, while Section 4 shows preliminary results of our system; we conclude with Section 5.

2. Related Work

Motion capture is not a new concept. The basics for camera calibration have been around for nearly three decades. Prior to that, photogrammetry had been used in a range of fields, from entertainment to medicine to robotics [5]. Vicon and Meta Motion are two of the larger cross-discipline motion capture companies.

The optical motion capture systems provided by Vicon and Meta Motion are all incredible products. However, the price for even the simplest of these optical systems is still in excess of $25,000 US [14] [16]. Faced with such financial odds, most independents and smaller studios resort to key-framing animation by hand or renting time at a motion capture studio.

In contrast to these precision optical systems, most animators have a far more basic yet broader need for motion capture. This solution introduces a portable, simple, inexpensive motion capture system to the masses.

The higher-end optical tracking systems use pulse-driven IR LEDs or reflective IR markers along with cameras equipped with IR lens filters. While this technique is an elegant solution for removing background noise, it restrains the actors to the stage in the studio. In contrast to this approach, our solution embraces the visible spectrum while minimizing any necessary environmental control.

The markers designed for our solution are similar in the regards that they are either reflective or optically active. High-intensity directional RGB LEDs replace the IR LEDs and reflective adhesive fluorescent tags replace the reflective IR markers.

One benefit of having an immobile motion capture studio is the minimal need to recalibrate the cameras. In our solution, the extrinsic parameters must be recalculated each time the cameras are repositioned. This is easily accomplished with a portable calibration pattern.

Overall, it seems the specialization of motion capture equipment and companies maintains a glass ceiling over those independent and smaller production houses.

3. Practical Motion Capture

As mentioned previously, our motion capture system includes several stages: camera calibration, marker selection and placement, and tracking, including post-processing to convert filmed markers to 3D co-ordinates. Each subsection describes our approach in detail.

3.1 Camera Calibration

Determining the intrinsic properties of each digital video camera in a motion capture system is critical, as the specific measurement of these properties determines the degree of precision attainable by the setup. These values include focal length, principal point, skew, and distortion [3]. The physical properties of the lens and the placement of the CCD can cause the image to be distorted (e.g., barrel or pincushion distortion), which of course, is highly undesirable for our purposes.

To determine the intrinsic properties of a camera, we use the standard technique of taking shots of a checkerboard pattern in multiple orientations. These images can be fed directly in the freely available Camera Calibration Toolbox in MATLAB for computing the intrinsic parameters of the camera [2] [4] (see Figure 1). (This program has also been implemented in C and is available within the Open Source Computer Vision library distributed by Intel [7].) Since this information does not change for an individual camera, it may be saved for future use.

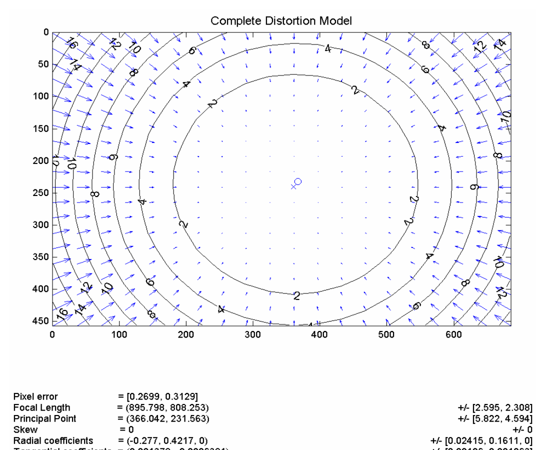

Figure 1: MATLAB – Intrinsic Radial Distortion

Extrinsic calibration is used to determine a camera's location and orientation within the environment and therefore must be performed each time the camera is moved or modified. Fortunately, the MATLAB Toolbox mentioned above can also calculate the extrinsic properties if the previously calculated intrinsic properties are available. Here, we once again use a checkerboard mat placed in the target environment to provide a reference pattern (see Figure 2). The squares within the image are all 30mm by 30mm and are within the field of view of all cameras.

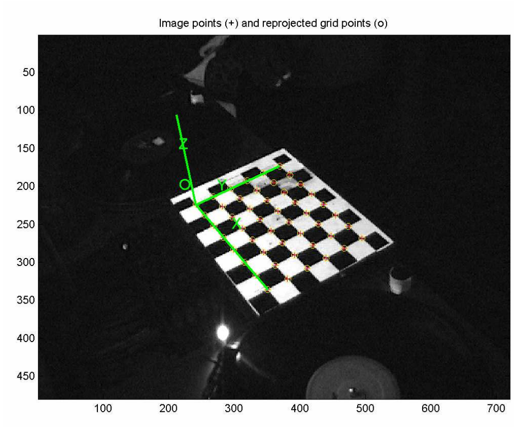

Figure 2: Calibration Mat

Of particular note are the focal length, principal point location, and the radial and tangential distortion coefficients [3]. Using this information, it is possible to correct tracked markers for radial and tangential distortion introduced by the camera lens.

3.2 Marker Selection and Placement

In creating a marker solution, several technical issues must be addressed: the type of marker used, including the type of power pack if necessary, and the integration of this hardware in a practical body suit.

We originally chose an optical system due to its flexibility and non-inhibitive characteristics. We also decided to use an active optical system, rather than a passive one, to allow the system to function in a broader range of environments. Once these decisions were made, we then needed to determine the best type of marker and its associated power source, in terms of portability and cost.

A widely used method for active optical markers incorporates LED's. The two main issues that arise when using LED's are their candela rating and the viewing angle. The candela is a metric unit of measurement equal to the amount of light emitted through a solid angle by the source. The Ultra Bright White LED's used in this solution have a rating of 8000-9000 mcd, which is good, but the viewing angle is a fairly narrow 60 degrees, which is problematic when trying to keep LED's in constant view during an actor's performance.

To contend with this limited viewing angle, we initially tried clustering several LED's to improve their visibility; however, this technique is problematic in that the centroid of the marker, which is needed to determine an accurate marker location, is difficult to detect. Figure 3 shows an LED cluster with five lights.

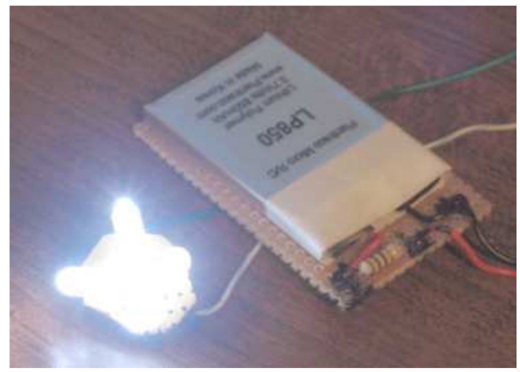

Figure 3: LED Cluster

This problem brought about the final design change for the markers. To reduce the number of LED's, while maintaining a larger viewing angle in the single marker, some sort of light refraction is necessary. Figure 4 shows a red refracting lens with a viewing angle of 180^{o}. The lenses used in the actual construction of our markers are clear.

Figure 4: Refracting LED Lens

The introduction of a refracting lens for each marker eliminates the need for clustering, and consequently, reduces the power needs for the body suit. In constructing a power solution, we consulted the large amount of research that has been done relating to the power/weight relationship of batteries. Figure 5 is an example of the cells used in this solution. Its dimensions are 50mm x 32mm x 5mm and its weight with the microconnector is 16.9 grams. The small size and specific energy of these cells make them ideal for mounting on hardware and, for this purpose, a person. Their small size also reduces any restrictions in movement that a larger battery set could impose upon an actor. Finally, these batteries do not suffer from recharge memory, as alkaline and other batteries do [9].

Figure 5: Lithium Polymer Cell

The body suit is a simple construct consisting of a single pair of thermal underwear. The wiring for each marker is run through a small fabric channel sewn onto the interior of the clothing. This approach allows the actor to move unrestrained and allows the wiring to shift depending upon the movement. Eyelets are attached to the exterior of the body suit at predetermined locations and a marker is attached to each. The power sources are grouped together in a small housing and attached to the back of the actor. The housing is of a general construction and simply allows the power source to be removed so the batteries may be easily recharged. An image of this low-cost body suit is shown in Figure 7(a).

3.3 Tracking

One of the larger problems with portable motion capture is finding an effective way to track markers in any myriad of lighting situations. The design of an active optical marker system was a key step in this multi-step process. With optical markers whose spectrum is of a predetermined value, filters may be defined within Shake to filter within any scene. This method is similar to one used by a number of companies which utilize active and reflective IR markers and IR camera lenses. However, the LED's selected for this project lie within the RGB spectrum.

Since the full visible spectrum is being utilized in this solution, Apple Shake is utilized heavily to modify the footage and track each marker. Adobe After Effects also contains a robust tracking tool and similar video compositing tools. Intel also offers an image processing library as well as supporting the open source computer vision library, OpenCV [6] [7] [13].

There are a number of advantages for this RGB range. It would allow the marker selection to be specifically tailored for each location. Also, it would permit the filming crew to "eye" each setup before and during each shoot. This spectrum information allows Shake to quickly extract the markers from what otherwise may be a busy scene. The process is similar in theory to using IR markers and camera filters within a controlled studio.

Once the Shake processing has been performed, we must tackle the problem of tracking multiple markers from multiple viewing angles. There are two main problems with this approach. The first relates to actually calculating the intersection location. In a perfect setup, two rays projected from their respective cameras towards a tracker marker would intersect at the tracking location. However, with cameras and software of less precision the error between the projected rays and their actual world location increases [5].

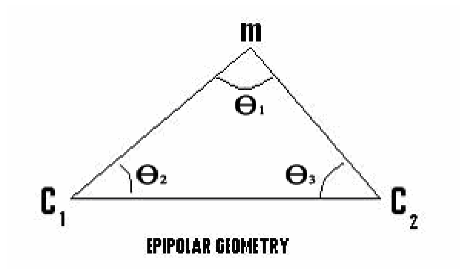

Figure 6: Epipolar Geometry

The ability of the software and hardware to determine this point is normally the proverbial yardstick by which such software is measured. Some approaches, such as the least-mean squares method, find the shortest line between each correlating pair of rays and uses the center of this line as the calculated intersection. For this solution, the method involving an epipolar base line between cameras simplifies the situation [3]. It also provides a method to differentiate between unnamed tracked objects.

If each camera is connected by a simple baseline, the intersecting rays from the cameras will form a triangle. If a camera tracks *n* objects in a frame, an *nxn* table of angles between the tracked marker location and epipolar base line is produced.

The epipolar plane is defined by an object point (*m*) and each camera's center of projection (C_1, C_2). Given a pair of cameras, this relationship forms a single triangle [11]. By using this geometric relationship, it is relatively easy to estimate the location of the object point. Figure 6 exhibits an example of this relationship.

This epipolar approach is especially useful as it allows for an initial estimation of the coordinate location. By calculating the two angles from the epipolar line and the marker, it is easy to differentiate between multiple markers within the same view volume. Including more cameras introduces greater accuracy in both coordinate calculation and marker labeling.

After triangulating the marker location, the generic labels are attached to each marker for future reconstruction. By giving labels to each marker, a skeleton may then be attached and used to calculate missing or occluded markers. Labeling may also be used to correct markers mislabeled due to proximity during tracking.

The issue of occluded markers is greatly dependent upon which tracking and reconstruction technique is used. There are several software methods available to reconstruct missing markers. Some methods search for contextual clues while other programmers utilize IK (inverse kinematic) chains to predict missing markers. Other companies take a less software-intensive approach and inundate the acting area with cameras to minimize occlusion. Independent of the approach used to track markers, there are certain issues pertaining to marker placement and the human body which must be resolved. These issues relate to the degrees of freedom specific to each joint in the human body [10].

There are number of ways to place markers on a human body for motion capture [8]. This solution utilized 11 markers but was designed for up to twenty. By using these methods to triangulate each marker location, we are ready to apply it to a virtual character.

4. Results

The motion capture process described above has been used in several different ways to animate various types of motion. In this section, we show how we applied this motion to a humanoid figure.

The image shown in Figure 7(a) is an arbitrary scene chosen to demonstrate the capabilities of our motion capture system. Note that no special lighting or background is needed, though a consistent background color may be easier to process than one with more detail (noise). Before this footage was shot, the cameras participating in the motion capture were calibrated using the portable checkerboard method described in Section 3, and all intrinsic and extrinsic parameters were recorded for later use.

The actor in the shot is wearing the low-cost marker suit with 8000-9000 mcd LED's. Note that some LED's, such as the one on the actor's left elbow, are not as bright as others, but are not required to point directly at the camera in order to be visible for tracking. This image is one of three perspectives shots that will be used to track the markers in this frame.

The image depicted in Figure 7(b) shows the same frame after processing with Shake. The small LEDs are picked up easily by the software. (As mentioned previously, other off-the-shelf or proprietary software would also suffice for this phase.) These processed frames will be used instead of the original footage for tracking markers.

Figure 8 shows the nine markers and their tracked coordinates over a two-second span. This processing is performed for all three camera perspectives for this frame. The information gathered from all three cameras is used to triangulate the location of each marker.

Once the data is collected and stored, it can be applied to virtual characters in a 3D animation and rendering package such as Maya. Of course, the skeleton of the virtual character must be created with the motion capture setup in mind such that marker locations calculated from the actor align with the joints of the virtual character. Figures 7(c) and 7(d) shows a simple rigged character in Maya in the position dictated by the motion capture data. This process is applied to sequences of images to fully animate the character.

5. Conclusion

The motion capture process described above has been used for several different projects to capture various types of motion. These motions include: a DJ interacting with a pair of turntables, an artist per-

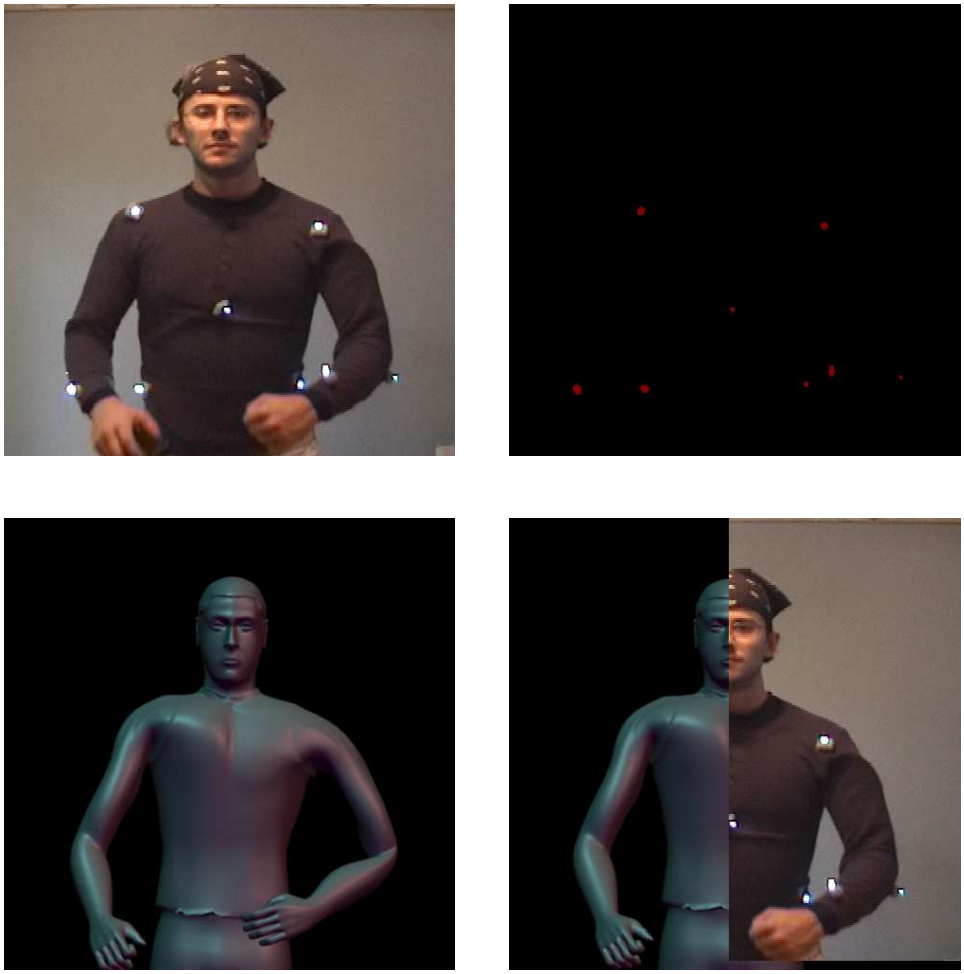

Figure 7: (a) original frame; (b) marker extraction; (c) virtual figure with motion capture data applied; (d) split image

forming Hawaiian fire spinning (or Poi), and actors exhibiting both full-body and facial motion.

The design and application of a portable calibration pattern for extrinsic camera calibration was a significant step in making this solution portable. The design of a lightweight non-tethered active-optical bodysuit allowed for the full body motion capture of an actor. By using a variation of active optical and reflective fluorescent markers in these situations, the markers were easily extracted from each scene. The added flexibility afforded by Shake's node structure allows for a myriad of filming conditions and subjects.

The biggest problems we faced stem from the cameras utilized: consumer-level Sony Handy-cams with maximum frame rate of 30 fps, small CCD's, and limited color management. Overall, however, the entire solution is quite inexpensive, not to mention relatively easy for a small group to implement. One of the greatest benefits, however, is having a solution which is easily transported between recording locations.

References

[1] B. Bodenheimer, C. Rose, S. Rosenthal, and J. Pella, "The Process of Motion Capture: Dealing with Data." *8th EUROGRAPHICS Int. Workshop on Computer Animation and Simulation*. '97, Budapest, Hungary, D. Thalmann and M. van de Panne eds., Springer-Verlag View, pp. 3-18.

[2] Camera Calibration Toolbox for Matlab, www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html, January 2004

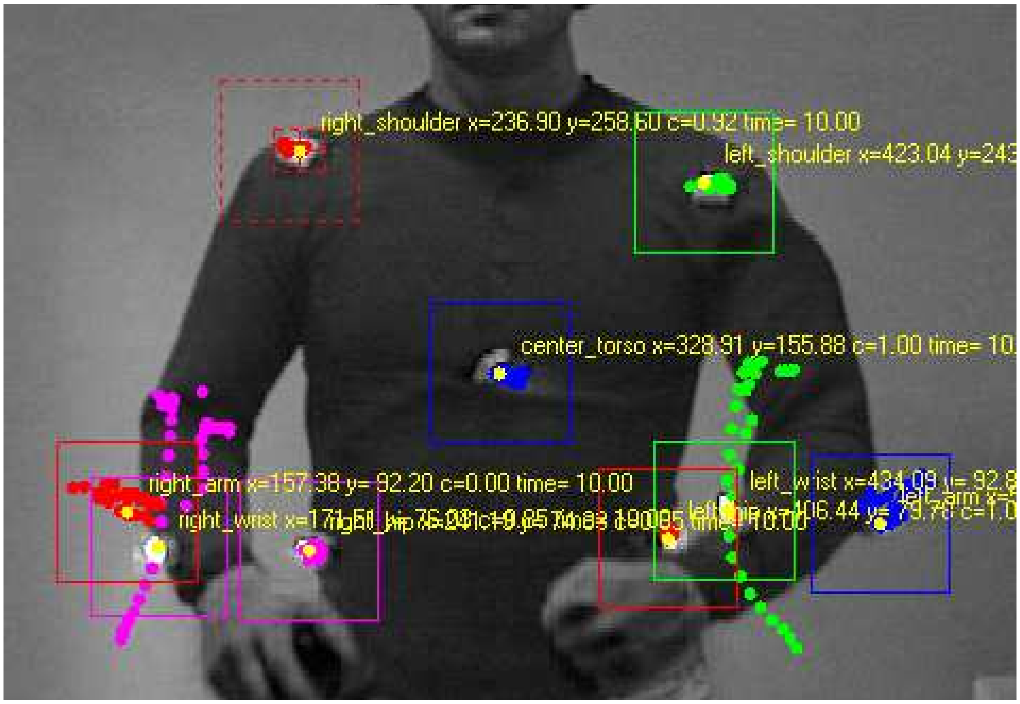

Figure 8: Tracking data in Shake

[3] "CVonline: On-Line Compendium of Computer Vision." R. Fisher (ed). homepages.inf.ed.ac.uk/rbf/CVonline/ March 2004.

[4] L. Dorst, S. Mann, and T. Bouma, "GABLE: A MAT-LAB Tutorial for Geometric Algebra," www.cgl.uwaterloo.ca/~smann/GABLE/, May 2004

[5] B. Horn. *Robot Vision*. MIT Press, 1986.

[6] Intel. IPO – Intel image processing library. www.intel.com/support/performancetools/ February 2004.

[7] Intel. Open source computer vision library. www.sourceforge.net/projects/opencvlibrary/. February 2004.

[8] I. V. Kerlow. *The Art of 3-D Computer Animation and Imaging.* John Wiley & Sons, Inc. 2000

[9] Lithium Battery Info. RCDrags.com, Forum Index, R/C Electric Discussion. August 2005. www.rcdrags.com/forum/viewtopic.php?t=601.

[10] R. Maiocchi, "3D Character Animation Using Motion Capture," in Advanced Interactive Animation, Thalmann & Magnen at-Thalmann (eds.), Prentice-Hall. Summer 1996.

[11] G. Olague, R. Mohr, S. Venkatesh, and B. C. Lovell, "Optimal camera placement to obtain accurate 3D point positions," in Proceedings. Fourteenth International Conference on Pattern Recognition, (Cat. No.98EX170), A. K. Jain, Ed. Los Alamitos, CA, USA: IEEE Comput. Soc, 1998 pages. 8-10 vol.1.

[12] R. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision" in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '86), pages 364-374, June 1986.

[13] C. Urmson, D. Dennedy, D. Douxchamps. G. Peters, O. Ronneberger, T. Evers. libdc1394 – open source IEEE1394 library. sourceforge.net/projects/libdc1394, February 2004.

[14] ViconPeak. www.vicon.com. April 2005.

[15] S. Wachter, H.-H. Nagel "Tracking of persons in Monocular Image Sequences" Proceedings, IEEE Nonrigid and Articulated Motion Workshop, June 16, 1997, San Juan, Puerto Rico

[16] Meta Motion. www.metamotion.com. April 2005