

# Software and Methods for Motion Capture and Tracking in Animation

**J Condell**

School of Computing and Intelligent Systems,  
Faculty of Engineering,  
University of Ulster at Magee College,  
Northland Road, Londonderry,  
Co. Londonderry, Northern Ireland

**G Moore, J Moore**

School of Computing and Mathematics,  
Faculty of Engineering,  
University of Ulster at Jordanstown,  
Shore Road, Newtownabbey,  
Co. Antrim, Northern Ireland

## Abstract

*This extended abstract details previous methods for motion tracking and capture in 3D animation and in particular that of hand motion tracking and capture. Our research aims to enable gesture capture with interpretation of the captured gestures and control of the target 3D animation software. This stage of the project involves the development and testing of a motion analysis system. A motion analysis system is being built from algorithms recently developed. We review current software and research methods available in this area and describe our work-in-progress.*

*Motion capture is a technique of digitally recording the movements of real entities, usually humans. It was originally developed as an analysis tool in biomechanics research, but has grown increasingly important as a source of motion data for computer animation. In this context it has been widely used for both cinema and video games. Hand motion capture and tracking in particular has received a lot of attention because of its critical role in the design of new Human Computer Interaction methods and gesture analysis. One of the main difficulties is the capture of human hand motion.*

**Keywords:** Animation, Hand tracking and analysis, Matchmoving, Motion capture.

## 1. Introduction

This extended abstract looks at various existing motion tracking and capture systems currently in use. Some diverse examples of motion capture being used in commercial applications are given. EyeToy® [8] is a motion sensitive USB camera that plugs into a play station 2. It uses your arms, legs or head to control a computer game. Another example is that of Annosoft [7] which provide automatic lip-sync products for games, multimedia, and video production, with a singular focus on delivering high quality speech products. Their system takes as input a wave or mp3 file and allows you to have text-less lip-sync or lip-sync with text. It moves the 3D models mouth automatically and outputs a file in '.anno' format which can then be used in 3D Studio Max using a plug-in.

In current work we are investigating tracking and capturing hand motion for animation. We propose the use of optical flow techniques in a vision-based system running in real-time.

Section 2 looks at the two main methods for capturing motion data: optical systems and magnetic systems. Section 3 details specific software and tools currently available for facial motion analysis. Section 4 continues by reviewing matchmoving in animation. In Section 5 we specifically focus on research methods developed for hand motion capture and tracking for applications such as hand pose estimation for HCI. Section 6 details our current work and where it sits within the literature.

## 2. Techniques for Capturing Motion Data

There are two main systems used for capturing motion data: optical systems and magnetic systems.

### 2.1. Optical Systems

Although there are many different systems for capturing motion data, one technique contains optical systems. These systems employ photogrammetry to establish the position of an object in 3D space based on its observed location within the 2D fields of a number of cameras. These systems produce data with 3 degrees of freedom for each marker, and rotational information must be inferred from the relative orientation of several markers.

For optical systems, the most common approach is to use passive reflective markers and to identify each marker from its relative location, possibly with the aid of kinematic constraints and predictive gap-filling algorithms.

A related technique ‘matchmoving’ can derive 3D camera movement from a single 2D image sequence without the use of photogrammetry. The term is used loosely to refer to several different ways of extracting motion information from a motion picture, particularly camera movement. Matchmoving is related to rotoscoping and photogrammetry and is sometimes referred to as motion tracking.

### 2.2. Magnetic Systems

Another technique for motion capture is the use of magnetic systems, which directly indicate the position and orientation of the sensors with respect to a transmitter. Magnetic systems use a centrally located transmitter and sensors that relay position and orientation in a measured space to capture motion. The sensors are able to measure their spatial relationship to the transmitter because they are immersed in an electromagnetic field.

Since the sensor output has 6 degrees of freedom, useful results can be obtained with a much smaller number of sensors than you would require with markers in an optical system. An advantage that magnetic systems have over optical systems is that the markers cannot be occluded, at least not in the way that they are using optical systems. Magnetic systems are also much cheaper than optical systems.

The major restrictions are that the response is quite nonlinear, especially near the edges of the capture area, and that the wiring from the sensors tends to preclude extreme movements on the part of the

performers. Also, the capture volumes for magnetic systems are dramatically smaller in size than they are for optical systems. Optical motion capture systems offer higher accuracy at higher sampling speeds than Electro-Magnetic systems and give greater freedom to the performers.

## 3. Facial Motion Capture

There are several systems and techniques for facial animation. Facial motion capture is challenging due to the subtle expressions possible from small movements of the eyes and lips, requiring even greater resolution and fidelity. Some use a special helmet with a video camera or an infra red camera and small reflective markers on the actors face.

1. **Marker-based systems** apply 10 to 100 markers to the actors face and track the marker movement with high resolution cameras.
2. **Markerless technologies** use features of the face such as nostrils, corners of lips and eyes, and wrinkles and track them.

There also exist video-analysis facial systems, voice recognition systems and text recognition systems. MOTEK offers two techniques for facial animation. The first is using VICON motion capture system, where optical markers are being attached to the actors face and the remainder is done using the same techniques as in full body motion capture session. The second system is a customized system which uses a special helmet with a lipstick/finger video camera and customized video tracking software (which does not need any markers or sensors) developed by U. K. based Image-Metrics [1]. Image Metrics specialize in creating unique image recognition based visual effects and complete facial reanimation using image understanding technology [1] – employing a wide range of bespoke technologies across Movies, TV, Computer and Video Games, Toys and Mobile Devices including photo-realistic manipulation of previously recorded footage.

The core of the patent technology is a set of software algorithms for analyzing and interpreting the content of video footage. The dedicated software first analyzes the original footage and employs a sophisticated automated reasoning to build a mathematical model of their 3D structure and surface textures. The result of this analysis is a set of mathematical structures that allow faces to be deformed, textured lit and manipulated into new

configurations to a level of next generation photo-realism. Image Metrics' technology also performs the same analysis of the footage of the new actors and then produces another set of mathematical models representing their 3D structure and face deformation. Image Metrics' patent technology called retargeting allows the performance of the actor. Every single frame of the actor's performance is analyzed in sequence and every deformation is applied to characters in the original footage. They create a visual effect that adds total photo-realistic manipulation and reanimation to virtually any new or recorded historic performance.

The Robotics Institute in CMU [2] has developed a variety of efficient real-time Active Appearance Models (AAMs) fitting algorithms. They initially developed an analytically-derived gradient-descent algorithm, based on their "inverse compositional" extension to the infamous optical flow techniques by Lucas and Kanade [3]. Compared to previous numerical algorithms, they showed their algorithm to be both more robust and faster. They also extended their 2D algorithm to fit "Combined 2D+3D Active Appearance Models," [4] an extension of an AAM that has both a 2D and a 3D shape model, thereby having the benefits of both. They claimed their method was even faster than the 2D algorithm because less iterations were required per frame. This speed-up illustrated the more constrained nature of fitting a 3D model.

Recently they also proposed a new extension of AAMs to multiple images [5] - the Coupled-View AAM. Coupled-View AAMs model the 2D shape and appearance of a face in two or more views simultaneously. The major limitation of Coupled-View AAMs, however, is that they are specific to a particular set of cameras, both in geometry and the photometric responses.

FaceLab™ [6] provides head-pose, gaze direction and eyelid closure tracking. It has an immediate and far-reaching impact in the realm of transportation safety and active information awareness systems. FaceLab™ is one of the most advanced marker-less motion capture systems on the market.

## 4. MatchMoving

"Camera tracking", "Matchmoving" or "3-D Tracking" is the process of analyzing a video clip or film shot to determine where in world-space the camera went, what its field of view was, and where parts of the set were. This is done by extrapolating 3D

data from the original 2D imagery. It is primarily used to track the movement of a camera through a shot so that a virtual camera move can be reproduced i.e. to render 3D objects, scenes and special effects with the same camera information. This allows the real scene to be matched with virtual creations and allows seamless compositing of the two scenes. There are many examples of matchmovie tools. A brief overview of a number of these is now presented:

### 4.1 Matchmovie Tools

*Voodoo* [9] (non-commercial) uses an estimation algorithm to give a full automatic and robust solution to estimate camera parameters for long video sequences. The estimated parameters can be exported to 3D animation packages. The method consists of four processing steps: automatic detection of feature points; automatic correspondence analysis; outlier elimination and estimation of the camera parameters

*Icarus* (non-commercial) is now defunct but apparently still used). *PixelFarm PFTrack* is the commercial reincarnation of *Icarus* [10]. They offer software for tracking QuickTime and AVI movies to produce 3D camera information that can be exported to your favourite 3D system or effects package. They can remove lens distortion, remove unwanted camera movement, and stitch footage together. They also use optical flow data to speed up or slow down a shot and to remove blur.

*RealViz Match Mover* [11] provides feature tracking. It is 3D tracking software which automatically or manually extracts 3D camera data with multiple objects motion from video or film sequences. The software can create slow motion or speed-up sequences.

*Ssontech SynthEyes* [12] is an automatic and supervised camera tracking and matchmoving system. It provides matchmoving the first shot in under a minute, exported to most major animation and compositing packages. *SynthEyes* can be used for animated character insertion, virtual set extension, accident reconstruction, architectural previews and virtual product placement.

*Scienc.D.Visions 3DEqualizer* [13] is 3D matchmover software which generates special effects for commercials, games and feature films. It incorporates motion-tracking features supported by mathematical algorithms with a user interface. *3D-Equalizer* enables the user to reconstruct precise 3D camera and object motion paths out of any type of live action footage. The range of platform-specific export capabilities is expanding, as is the feature packet

delivered by 3D-Equalizer's tracking engine. 3D-Equalizer V3 can acquire motion capture data. Multiple cameras record a moving object simultaneously from different angles. These cameras can be either static or moving ones. In some situations a camera can be replaced by a simple mirror, so that motion capturing with a single camera is possible.

*2D3 Boujou* [14] claims to be the world's first automatic matchmoving application. 2D3 offer 2 packages with a suite of tools that enable you to automatically feature track footage.

*Simi Reality Motion Systems* [15] offer software for motion capture, automatic tracking, coaching, athlete feedback as well as a notational system and motion analysis for scientific and educational purposes. Simi Motion performs 2D or 3D motion capture and analysis. Capturing of the movement is not time-limited and after digitization all data can be edited and visualized in many ways. 3D coordinates can be synchronized with data from other devices and exported in various formats. Simi MotionCap 3D is software created for entertainment applications like 3D animation. It enables real motion sequence recording from several perspectives synchronously, capturing them and processing the emerging data. The 3D movement data can be exported to common 3D applications. Simi MatchiX is image processing software for automatic markerless tracking. The pattern matching algorithm can be utilized with video clips or still image sequences and automatically tracks user-defined patterns of different sizes. Thus, the system provides all necessary translation, rotation and time information about the markers. Simi MotionTwin provides video-based motion analysis using static kinematics.

## 5. Hand Tracking using Motion

The two main methods for hand tracking are appearance-based methods and model-based methods. Appearance-based methods generally establish a mapping between the image feature space and the hand configuration space. Model-based methods are generally deformable hand shape models fitted with statistical models. Kinematic models are also used. Recently more tracking-by-detection methods have emerged which merge these two categories by searching exhaustive databases. In this section we will detail various research methods developed for hand motion capture and tracking.

3D hand tracking has great potential as a tool for better human-computer interaction [16]. Tracking

hands, in particular articulated finger motion, is a challenging problem because the motion exhibits many degrees of freedom. Self-occlusion can cause problems with hand tracking, as can tracking in cluttered backgrounds, and automatic tracker initialization. 3D tracking differs from gesture recognition, where there is a limited set of hand poses which need to be recognized. Stenger [16] investigated model-based hand tracking using a hierarchical Bayesian filter. Essentially the tracker is a tree-based filter, which approximates the optimal Bayesian filtering equations. The 3D geometric hand model is built from truncated quadrics and its contours can be projected into the image plane while handling self-occlusion. Articulated hand motion was learned from training data collected with a data glove, leading to a lower dimensional representation of finger motion. Edge orientation and skin colour information was used, making the matching more robust in cluttered backgrounds.

Benoit and Ferrie [17] used a near real-time optical flow algorithm to compute motion using region-based matching techniques. They carried out various experiments on standard image sequences. They also demonstrated how their algorithms would be useful for 3D shape recovery, particularly where an object is held and waved in front of a camera by a hand. Upcoming flow could be predicted which could be used in a practical machine vision application.

Dewaele et al. [18] tracked full hand motion from 3D points on the surface of the hand. They reconstructed and tracked these points using a set of cameras. They combined optical flow methods with 3D reconstruction at each time frame to capture the motion of the hand. Their hand motion model used animation techniques to represent the skin motion near joints.

Metaxas et al. [21] developed an articulated dynamic hand model driven by multiple cues including an extended optical flow constraint along with edges which permitted tracking of different hand motions. They used a probabilistic framework and showed the results of their algorithm applied to a single camera sequence.

Lee and Cohen [22] focused their research on the accurate detection and tracking of un-instrumented hands for assessing user performance in accomplishing a task. They automatically tracked hand motion and recognized the corresponding gestures. Their objective was to augment or replace the mouse and keyboard paradigm with functionalities relying on natural hand motion for driving a specific

application. Their approach used inter-finger constraints and global motion constraints to divide hand motion into global pose and individual finger motion.

A ‘Flocks of Features’ method was described by Kolsch and Turk [23] which tracked hands in live video combining optical flow image cues and a learned colour probability distribution. A hand gesture recognition system which integrates their Flocks of Features method into a vision-based interface is provided as open-source code.

Another method used for hand tracking has been eigen-dynamics analysis to learn the dynamics of natural hand motion from labeled sets of motion captured with a glove [24]. A Bayesian network was described and a hand tracking system implemented.

Heap and Hogg [25] developed a deformable point distribution model for 3D hand tracking, capturing training data semi-automatically from volume MRI via a physically-based model. Their method did not require the use of markers and showed how 2D information could be extracted to move and deform a 3D model. Their main limitations and problems were with occlusions.

### 5.1. Hand Pose Estimation

Hand motion capture and tracking is used widely in hand pose estimation systems (hand posture analysis). Hand posture analysis assists in developing new applications such as 3D hand gesture recognition. There are methods which use gloves and those which are vision-based systems. Vision-based systems can be classified into contour-based systems and model-based systems. The model-based methods again fall into two categories: those with markers and those without markers. A couple of methods are described here to show how hand tracking leads on to hand pose estimation.

Lin et al. [19] captured hand motion using a divide and conquer approach to estimate local and global hand motion. They also determined the hand pose using an Iterative Closed Point algorithm. A sequential Monte Carlo technique allowed them to efficiently track the finger motion. They claimed their approach was accurate and robust for natural hand movements.

3D hand pose estimation from a single image has been investigated by Athitsos and Sclaroff [20]. They treated it as a database indexing problem and generated a large database of synthetic views. Given the input image of a hand, the most similar images were retrieved and the pose parameters used as

estimates for the input pose image. They also used a Cyber Glove to monitor the angular motions of the palm and fingers. A limitation of this method is that it required very clean hand segmentation.

## 6. Current Work

There have been a number of attempts to develop more intuitive input devices, not all of them solely for the purposes of 3D modeling and performance related issues in commercial animation studios. General mouse-like devices are widely available. These are essentially input devices that allows data to be entered more efficiently but which add little value to the data over conventional input methods. An approach that provides richer data uses joystick and data glove devices to control digital models and physical puppets. However, such approaches still use physical interfaces and are generally large scale solutions unsuitable for animators to use at their workstations. There have also been attempts to develop systems that do not require a physical interface. However, at present, these tend to be focused on higher-level interactions than those required for modeling and animation. None of these approaches fully realise the potential to provide a small scale solution that in addition to increasing the efficiency of data input also adds value to the data being entered in order to improve the quality of the finished work and its efficient production. This is the problem that current work-in-progress is attempting to address.

Existing techniques for animating 3D computer characters can be time consuming and detailed in application, requiring the use of complex and often unintuitive user interfaces. This often results in the animation process having a negative effect on the nature of the finished animation. Current work-in-progress employs computer vision techniques to develop a prototype desktop product and associated animation process that will allow an animator to control character animation through the use of hand gestures. It is important to note that the hand gestures will form the basis of a performance capture system to facilitate a form of virtual puppeteering, rather than taking a motion capture approach. This should provide a softer, more intuitive, user interface for the animator that should improve the productivity of the animation workflow and the quality of the resulting animations.

The work is divided into two main stages: a simple hand-to-hand mapping (essentially real-time motion capture) and modal analysis. This development of the

work would repurpose the hand data to a limited number of abstracted values which would then alter the motion curves in a given 3D model. Different modes would need to be selected for hands, faces, walk, cycles etc. Per-user configuration would be provided to compensate for differing hand shapes and personal preference.

Small animation studios, which make up a substantial proportion of the industry, could benefit from an input device that not only has the potential to speed development and production but that could also help base their animations firmly on performance. For example, each action could be rehearsed and repeated until perfected and then the selected performance could be blended with other appropriate takes. If made available as a simple plug-in, it could be ported to work with most existing 3D packages.

It is well known that approaches that use gloves and other input devices can be expensive, cumbersome and difficult to use. Also the use of hand markings and the need for highly constrained environments are considered to be undesirable. Vision-based hand tracking is a cost-effective, affordable and non-invasive technique. Database and appearance-based approaches require a large amount of training data in order to achieve good results. Alternatively, model based approaches require a search in high dimensional spaces with up to more than 20 degrees of freedom. As with any vision system occlusions can also cause problems. Given that the current research is aimed at providing small animation studios with a new tool, a further constraint is that the resulting system should use affordable and readily accessible components and a simple configuration. To this end it is anticipated that the system will use a single low-cost camera for input and will run in real-time on a workstation with a typical specification.

In current work we are investigating tracking and capturing hand motion for application in 3D animation. The goal is to produce a prototype camera based desktop gesture capture system to capture hand gestures and interpret them in order to control the animation of 3D character models within industry standard animation software. Figure 1 shows an overview of the configuration of the proposed system.

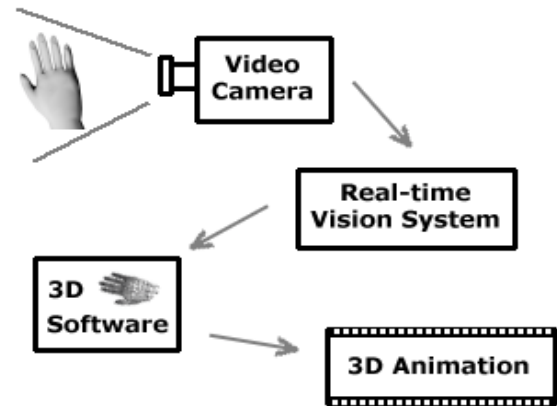


Figure 1: Overview of current system

To this end we are developing a vision system that uses optical flow techniques, particularly recently developed motion estimation techniques [18], to analyse live video from a single camera. This modified version of the “Horn and Schunck” motion estimation technique [18] shows improvements in computational efficiency by “focusing on those regions of the image sequence in which motion is detected to be occurring”. The main improvements achieved through the algorithms developed through two-dimensional adaptive approaches will be exploited in current work.

The output from the vision system will be transferred in real-time to an industry standard animation package within which it will be used to control an appropriately rigged 3D model. The use of real-time techniques and industry standard animation software will facilitate integration of the system into existing workflows in order to keep costs low while at the same time show improved quality and productivity.

## 7. References

- [1] Image Metrics: <http://www.image-metrics.com/>
- [2] Real-Time AAM Fitting Algorithms by the Robotics Institute, CMU: [http://www.ri.cmu.edu/projects/project\\_448.html](http://www.ri.cmu.edu/projects/project_448.html)
- [3] Lucas and Kanade, 1981. “An Iterative Image Registration Technique with an Application to Stereo Vision”. *Proceedings of the 7<sup>th</sup> International Joint Conference on Artificial Intelligence (IJCAI '81)*, pp. 674-679.
- [4] Matthews and Baker, 2004. “Active Appearance Models Revisited”. *International Journal of*

*Computer Vision (IJCV)*, Kluwer Academic Publishers, vol.60, no.2, pp.135 - 164.

[5] Koterba et al., 2005. "Multi-View AAM Fitting and Camera Calibration". *Proceedings of the International Conference on Computer Vision (ICCV)*, pp.511-518.

[6] FaceLab™:

<http://www.seeingmachines.com/facelab.htm>

[7] Annosoft: <http://www.annosoft.com/>

[8] EyeToy®: <http://www.eyetoy.com/index.asp>

[9] Voodoo:

<http://www.digilab.uni-hannover.de/docs/manual.html>

[10] PixelFarm PFTrack:

<http://www.thepixelfarm.co.uk/>

[11] RealViz Match Mover: <http://www.realviz.com/>

[12] Ssontech SynthEyes: <http://www.ssonotech.com/>

[13] Scienc.D.Visions 3DEqualizer:

<http://www.3dequalizer.com>

[14] 2D3 Boujou: <http://www.2d3.com/jsp/index.jsp>

[15] Simi Reality Motion Systems:

<http://www.simi.com/en/>

[16] Stenger, 2004. "Model-Based Hand Tracking Using a Hierarchical Bayesian Filter". *PhD Thesis*, Department of Engineering, University of Cambridge.

[17] Benoit and Ferrie, 1996. "Monocular Optical Flow for Real-Time Vision Systems". *Proceedings of the 13<sup>th</sup> International Conference on Pattern Recognition (ICPR)*, vol. 1, pp.864-868.

[18] Dewaele et al, 2004. "Hand Motion from 3D Point Trajectories and a Smooth Surface Model". *Proceedings of the 8<sup>th</sup> European Conference on Computer Vision*, LNCS 3021, vol.I, pp.495-507.

[19] Lin et al., 2002. "Capturing Human Hand Motion in Image Sequences". *Proceedings of IEEE Workshop on Motion and Video Computing (WMVC02)*, pp.99-104.

[20] Athitsos and Sclaroff, 2003. "Estimating 3D Hand Pose from a Cluttered Image." *IEEE Computer Society, Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, vol.II, pp.432-439.

[21] Metaxas et al., 2003. "Using Multiple Cues for Hand Tracking and Model Refinement". *IEEE Computer Society, Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, vol.II, pp.443-450.

[22] Lee and Cohen, 2004. "3D Hands Reconstruction from Monocular View". *Proceedings of the International Conference on Pattern Recognition (ICPR)*, vol.III, pp.310.

[23] Kolsch and Turk, 2005. "Hand Tracking with Flocks of Features". *IEEE Computer Society,*

*Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, vol.2, pp. 1187.

[24] Zhou and Huang, 2003. "Tracking Articulated Hand Motion with Eigen Dynamics Analysis". *Proceedings of the 9<sup>th</sup> International Conference on Computer Vision (ICCV)*, vol.2, pp.1102.

[25] Heap and Hogg, 1996. "Towards 3D Hand Tracking using a Deformable Model". *Proceedings of the 2<sup>nd</sup> International Conference on Automatic Face and Gesture Recognition (FG '96)*, pp.140.

[26] Condell et al., 2005. "Adaptive Grid Refinement Procedures for Efficient Optical Flow Computation." *International Journal of Computer Vision (IJCV)*, Kluwer Academic Publishers, vol.61, no.1, pp.31-54.