EG UK Theory and Practice of Computer Graphics (2006)
M. McDerby, L. Lever (Editors)

Hand Motion Capture and Tracking in 3D Animation

J Condell

School of Computing and Intelligent Systems, Faculty of Engineering,
University of Ulster at Magee College, Northland Road, Londonderry, Northern Ireland
j.condell@ulster.ac.uk

G Moore, J Moore

School of Computing and Mathematics, Faculty of Engineering,
University of Ulster at Jordanstown, Shore Road, Newtownabbey, Northern Ireland
g.moore@ulster.ac.uk

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.

1. Introduction

This extended abstract looks at various existing motion tracking and capture systems currently in use. In current work we are investigating tracking and capturing hand motion for 3D 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 motion analysis, mainly facial motion analysis. Section 4 continues by reviewing camera tracking, matchmoving or 3D tracking 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. A conclusion is provided to the paper in Section 7.

2. Techniques for Capturing Motion Data

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

Magnetic systems for motion capture 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. 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.

Optical motion capture systems offer higher accuracy at higher sampling speeds then Electro-Magnetic systems and give greater freedom to the performers.

3. Facial Motion Capture

There are several systems and techniques for facial animation in particular. 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.

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

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].

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-decent 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.

FaceLab™ [4] 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 "3D Tracking" are processes 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 [5] (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

PixelFarm PFTrack [6] 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 [7] 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 [8] 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.

Sciene.D.Visions 3DEqualizer [9] 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.

2D3 Boujou [10] 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 [11] 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. 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 Optical Flow

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 improved human-computer interaction [12]. 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 [12] investigated model-based hand tracking using a hierarchical Bayesian filter. Articulated hand motion was learned from training data collected with a data glove, leading to a lower dimensional representation of finger motion.

Benoit and Ferrie [13] used a near real-time optical flow algorithm to compute motion using region-based matching techniques. They carried out various experiments on standard images sequences. Upcoming flow could be predicted which could be used in a practical machine vision application.

Dewaele et al. [14] tracked full hand motion from 3D points on the surface of the hand. They combined optical flow methods with 3D reconstruction at each time frame to capture the motion of the hand.

Lu et al. [15] 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 [16] 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 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 [17] which tracked hands in live video combining optical flow image cues and a learned colour probability distribution.

6. Proposed Project

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. This project 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.

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. However, such approaches do have their own constraints. 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 proposed 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.

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. The output from the vision system will then 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.

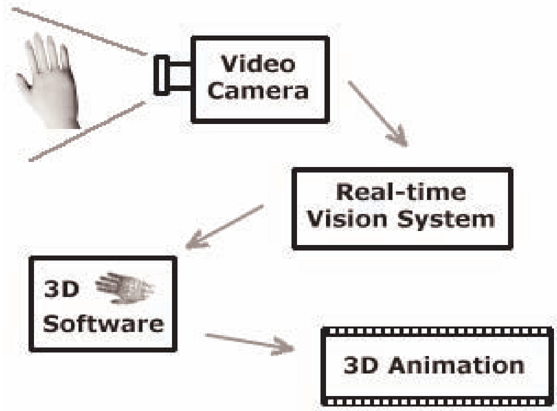

Figure 1: Overview of proposed system

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

3. B. Lucas and T. Kanade. “An Iterative Image Registration Technique with an Application to Stereo Vision”. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81),* 1981, pp. 674-679.

4.FaceLab™: http://www.seeingmachines.com/facelab.htm

5. Voodoo: http://www.digilab.uni-hannover.de/docs/manual.html

6. PixelFarm PFTrack: http://www.thepixelfarm.co.uk/

7. RealViz Match Mover: http://www.realviz.com/

8. Ssontech SynthEyes: http://www.ssontech.com/

9. Sciene.D.Visions 3DEqualizer:
http://www.3Dequalizer.com

10. 2D3 Boujou: http://www.2D3.com/jsp/index.jsp

11. Simi Reality Motion Systems:
http://www.simi.com/en/

12. B. D. R. Stenger. “Model-Based Hand Tracking Using a Hierarchical Bayesian Filter”. PhD Thesis, Department of Engineering, University of Cambridge, 2004.

13. S. M. Benoit and F. P. Ferrie. “Monocular Optical Flow for Real-Time Vision Systems". *Proceedings of the International Conference on Pattern Recognition (ICPR '96)*, 1996, vol. 1, pp.864-868.

14. G. Dewaele, F. Devernay and R. Horaud. “Hand Motion from 3D Point Trajectories and a Smooth Surface Model”. *Proceedings of the European Conference on Computer Vision (ECCV '04)*, 2004, LNCS 3021, vol.I, pp.495-507.

15. S. Lu, D. Metaxas and D. Samaras. “Using Multiple Cues for Hand Tracking and Model Refinement”. *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR '03),* IEEE Computer Society Press, 2003, vol. II, pp.443-450.

16. S. U. Lee and I. Cohen. “3D Hands Reconstruction from Monocular View”. *Proceedings of the International Conference on Pattern Recognition (ICPR '04)*, 2004, vol. III, pp.310.

17. M. Kolsch and M. Turk. “Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration”. *Proceedings of the Computer Vision and Pattern Recognition Workshop (CVPRW '04)*, 2004, Vol.10, pp. 158.

18. J.V. Condell, B. W. Scotney and P. J. Morrow. “Adaptive Grid Refinement Procedures for Efficient Optical Flow Computation.” *International Journal of Computer Vision (IJCV)*, Kluwer Academic Publishers, 2005, Vol .61, no. 1, pp.31-54.