Visual Tracking for Sports Applications

Andrew W. B. Smith and Brian C. Lovell
Intelligent Real-Time Imaging and Sensing Group, EMI
The School of Information Technology and Electrical Engineering
The University of Queensland
Brisbane, Qld 4072, Australia
{awbsmith, lovell}@itee.uq.edu.au

Abstract

Visual tracking of the human body has attracted increasing attention due to the potential to perform high volume low cost analyses of motions in a wide range of applications, including sports training, rehabilitation and security. In this paper we present the development of a visual tracking module for a system aimed to be used as an autonomous instructional aid for amateur golfers. Postural information is captured visually and fused with information from a golf swing analyser mat and both visual and audio feedback given based on the golfers mistakes. Results from the visual tracking module are presented.

1. Introduction

Visual tracking of human movement has attracted much attention due to the wide variety of applications which could be performed autonomously however currently need human interpretation. These applications include sports training, rehabilitation and security. Autonomous interpretation of human movement allows a much larger volume of analyses to be performed at a much reduced cost. Biometric analysis has already established itself as an effective training tool for athletes, although most techniques rely on the use of retro–reflective markers or magnetic sensors to be placed on an athlete before such analysis can be performed.

The aim of this project is the development of a system which uses visual cues to obtain a golfers postural information, and analyzes this with respect to a learned ideal motion. This data is then fused with information from a golf swing analyser mat which detects information about the club head which is infeasible to detect visually. Completely automated feedback can then be given based on differences between the athletes motions and the technically correct motion. Golf has been chosen as the sport in focus due to the limited movement of the player and the presence of an ideal motion. Smith and Lovell [16] give a more detailed description of the system and the swing analyser mat. In this paper we focus on the visual tracking module of this project. We provide some background literature and show results from the visual tracking module.

2. A Brief Overview of Tracking Algorithms

Algorithms to perform human tracking from multiple views can be thought of as being in two categories; deterministic and stochastic.

2.1. Deterministic Tracking Algorithms

Deterministic algorithms assume that the human body position can be uniquely determined at each point in time. Luck *et al.* [10, 9] and Small [15] adopt a deterministic approach where they construct a visual hull using shape from silhouette methods and fit a body model to it using a physics based fitting mechanism. Luck *et al.* [9] achieves tracking at 9Hz (each frame of video requires .11s to process) using 25mm^3 voxels and a 25 degree of freedom (DOF) human model in this manner. Mikic *et al.* [13] adopt a similar approach whereby they again form the visual hull from shape from silhouette methods however use an extended Kalman filter to fit the body model. They achieve tracking at 10Hz using 25mm^3 voxels and a 23 DOF human model.

The methods described above rely on background subtraction methods to produce an accurate volumetric hull. In the event of motion in the background, or some outdoor settings, background subtraction will not be sufficient to form an accurate visual hull. In theses cases it is not always possible to uniquely determine the body position from a practical feature set. Generally events like background clutter and occlusion prevent the body position from being uniquely determined at a given time.

2.2. Stochastic Tracking Algorithms

Stochastic algorithms do not rely on the body being uniquely determined at each point in time. Instead they assign probabilities to possible body positions and seek the most probable position. The Particle Filter, first used in visual tracking by Isard and Blake [7], was introduced to successfully track in the event of multi modal probability density functions (pdf). Particle Filters approximate a pdf by sampling from it. Predictions of the object position at the next time step are based on the probabilities of these samples (particles). In this way a particle filter can retain multiple hypotheses of the objects position. Deutscher *et al.* [2] improved the performance by adding annealing layers to the algorithm, allowing the pdf to be more extensively sampled from in regions of interest, generally the high probability regions. A further improvement was made by Deutscher *et al.* [3] by varying the amount of noise added to each particle during the sampling process, and introducing a crossover operator similar to that in genetic algorithms.

A problem for particle filters is that the higher dimensionality of the configuration space, the more particles required, and hence the higher the computational cost. MacCormick and Blake [11] showed that the number of particles required, N, can be found by

$$N \geq \frac{D_{min}}{\alpha^d} \tag{1}$$

where D_{min} and $\alpha \ll 1$ are constants, and d is the dimensionality of the search space. Deutscher *et al.* [3] report successful human tracking at 0.07Hz using a 29 DOF human model.

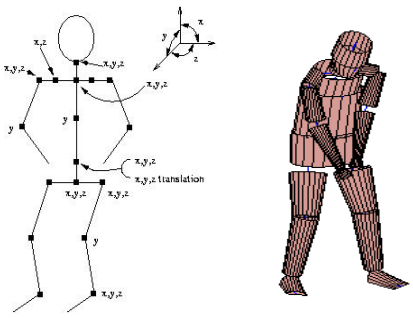
Sminchiescu and Triggs [14] present an alternative approach to stochastic tracking using the Covariance Scaled Sampling (CSS) algorithm. CSS propagates a multi-modal prior, essentially a mixture of Gaussians, and locally optimizes the new estimates such that they correspond to local minima in the posterior. Minima are sought as optimization involves minimizing a cost function as opposed to maximizing a pdf. During propagation, each Gaussian is sampled from according to the shape of the cost function, allowing sampling to be biased along the directions of most uncertainty. During optimization several samples may converge to the same local minima. Sampling in this way reduces the number of particles required for successful tracking as samples are better chosen to lie in regions of interest, and are optimized to reach minimas instead of randomly finding them as with particle filters. This method was primarily developed for monocular tracking, where the cost function is ill conditioned as approximately one third of joint variables are unobservable at each time instance. The key to using this approach is that the cost function is in some sense smooth, meaning local minima are not clustered together. To achieve this Sminchiescu and Triggs incorporate motion boundaries, intensity edge energy, optical flow and body model priors to form a robust cost function. They achieve monocular tracking of a 30 DOF human model at 0.0056Hz.

When the nature of the application allows for post processing of tracking results, a backwards optimization phase can be added to the stochastic tracking algorithms to improve results. Isard and Blake [8] present a framework for an output smoothing filter. The smoothing filter can be thought of as finding the Baum-Welch solution to the best path through a Hidden Markov Model, where the transitional probabilities are derived from a dynamic model. This smoothing filter provides a powerful tool when multiple hypotheses of the object position are present.

3. Modelling the golfer

Human tracking applications generally use about 30 degrees of freedom (DOFs) to model a person. These models are overly simplified for the task of tracking a human during a golf swing however. Currently we use 42 DOFs consisting of 3 translational and 39 rotational DOFs as shown in Figure 1(a). This a high dimensional space for he particle filter to search through, and the amount of particles hence computational time needed for the particle filter grows exponentially with the dimensionality of the space.

Our model model is constructed as a link list, where each link has a set of rotations and a surface modelled by a truncated elliptical conic, shown in Figure 1(b). A similar approach was used by Deutscher *et al* [2] and Goncalves *et al.* [5]. Sminchisescu *et al* [14] uses shape deformable super quadratic ellipsoids to model the surface, and Fua *et al.* [4] uses a summation of n three dimensional Gaussian density distribution known as metaballs. We use truncated elliptical conics as they are computationally cheaper and do not require any DOFs to model them.



(a) Degrees of Freedom (b) The Model Surface

Figure 1. Modelling the articulated body

Since we have an expectation about the possible postures

a golfer can take during the swing, a principal component analysis (PCA) could be used to reduce the dimensionality of the search space. In time this will be done, however currently only two golf swings have been manually annoted - as it is a time consuming process. Once tracking results have been obtained that are representative of all the possible postures of the golfer it is hoped a PCA can be performed to reduce the configuration space to around 20 dimensions. In the case of the golf swing, we know the hands must always be holding the club. This information could also be used to restrict the search space. Currently any configuration where the hands are more than a threshold distance apart are given a zero probability.

3.1. Dynamic Model

Due to the specific nature of the tracking in this case, a dynamic model can be used to improve the trackers performance. As mentioned above, only two swings have been manually annoted, each of which consists of 55 frames. Using a second order dynamic model in the 42 DOF search space, we have $2 \times 42 \times 55 = 4620$ equations with which to solve for $2 \times 42^2 + 42 = 3570$ variables. Due to the similarity between the two hand annoted swings, the dynamic model proved too powerful resulting in a near singular noise covariance matrix. To over come this, a PCA was performed to reduce the search space to 13 dimensions, giving a 25% variable to equation ratio. This dynamic model was then transformed back into the original 42 dimensional space, resulting in a practical noise covariance matrix.

4. Cameras

In this application we use Dragonfly cameras from Point Grey Research [17] . They synchronously capture 640×480 color images at 30 frames per second.

Since it is desirable to keep the system as small as possible, low focal length cameras are needed so the cameras can be placed as closely as possible to the golfer. This introduces radial distortion which we estimate using a technique described by Hartely and Zisserman [6], whereby parameters are chosen to make real world straight lines straight in the image. To choose the order of the radial correction function, Consistent Akaikes Information Criteria (CAIC) described by Bozdogan [1] is used. The results are shown in Table 1, with a second order model being used as it has the smallest CAIC value.

Projection matrices were determined from real world and image point correspondences, using the DLT algorithm with non-linear optimization described by Hartley and Zisserman.

Model Order	L2 Norm Error	CAIC Value
0	1695.114	Inf
1	84.9189	101.2427
2	5.4762	26.6286
3	5.3464	31.3275
4	5.2819	36.0917
5	5.2782	40.9165
6	5.2770	45.7441
7	5.2765	50.5722
8	5.2762	55.4005

Table 1. Model Order Selection

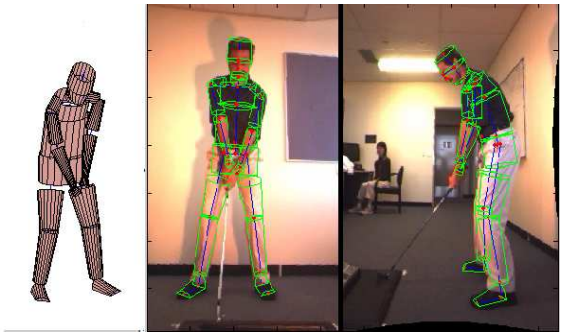
5. Results

The tracking results presented here are performed used the APF algorithm described in Section 2.2. The PAPF algorithm was not used due to its incompatibility with the output smoothing filter also described in Section 2.2, which was applied to our results. The body model was assumed known apriore, however background was assumed unknown during the tracking. An office environment was used purely for the convenience of capturing the footage and calibrating the cameras. The tracker was initialized by setting the model to an approximate golf address position and then using an APF to do a quick translational search.

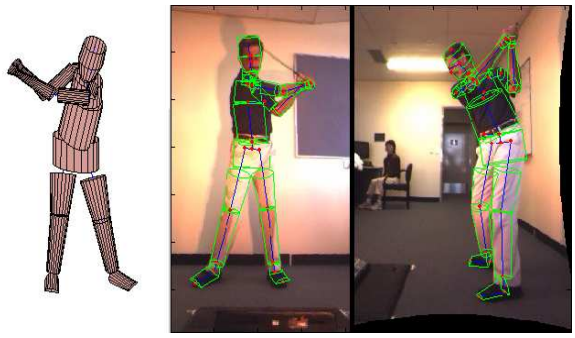
Observational probabilities were determined by casting measurement lines tangential to the projection of the link surfaces. Features along the measurement lines were found by high pass filtering the grey scale values along these lines, with features being points above a set threshold. Details of this method are given by MacCormick [12] . The probabilities for each measurement line were combined using a sum of squared differences approach, as used by Deutscher *et al.* [2]. Deutscher *et al.* [2] uses a different method to determine probabilities, they build an edge map for the image and assign probabilities based on the proximity of a sampled point to an edge from the edge map. We did not adopt this method as we assert the measurement line approach is more sensitive to low contrast features, such as exist between the left upper leg and the wall in Figure 2. We do concede however that our approach generally producing a less smooth pdf, i.e the pdf contains many more local maxima and so is more difficult to search.

Self occlusion models were used, with an added constraint that should the measurement line pass through another link the same color the measurement line was counted as occluded. This was done so if, for example, the upper arm was beside the torso features would not be expected between the two links.

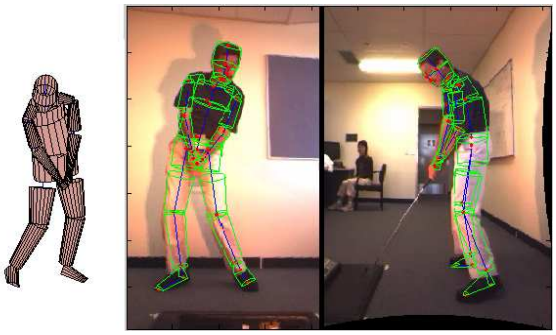
The edge probabilities were combined with color probabilities by adding their sum of squared differences. The color probabilities were determined by taking the interior



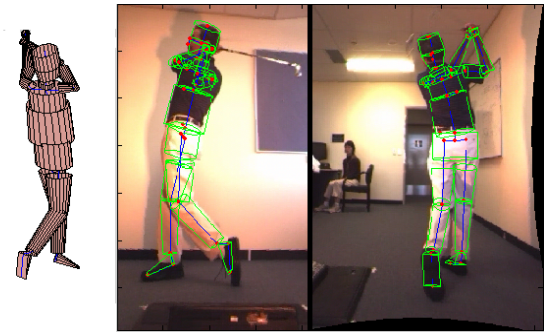
(a) Frame 1



(b) Frame 15



(c) Frame 33



(d) Frame 52

Figure 2. Results of Tracking at Selected Frames

most point on each measurement line, and comparing it to a known distribution of the links color.

Figure 2 shows tracking results at selected frames. Note the cameras are calibrated to act as mirrors as it is easier to give feedback in that manor. Each frame took approximately 25 minutes to process on a P3 833Mhz machine, with a MATLAB implementation of the APF. A video of the tracking results can be found at http://www.itee.uq.edu.au/˜iris/.

6. Conclusions and Future Work

Here we have shown that accurate tracking of the golfer during a standard golf swing is tractable without the need for background subtraction. The dynamic model proves a powerful tool for tracking in the high dimensional space used to model the golfer. Future work will include learning the body model from the video sequence, removing the color probability from the observational model as well as reducing the time required for tracking.

References

[1] H. Bozdogan. Model Selection and Akaikes Information Criterion (AIC):The General Theory and its Analytical Extensions. *Psychometrika*, 52(3):345-370, 1987.

[2] J. Deutscher, A. Blake and I. Reid. Articulated body motion capture by annealed particle filtering. *Proceedings of Computer Vision and Pattern Recognition Conference*, 2:126-133, 2000.

[3] J. Deutscher, A. J. Davison and I. Reid. Automatic Partitioning of High Dimensional Search Spaces associated with Articulated Body Motion Capture. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[4] P. Fua, A. Gruen, N. D'Apuzzo and R. Plankers. Markerless Full Body Shape and Motion Capture from Video Sequences. *International Archives of Photogrammetry and Remote Sensing*, 34(5):256-261, 2002.

[5] L. Goncalves, E. D. Bernado, E. Ursella, and P. Perona. Monocular Tracking of the Human arm in 3D. *ICCV95*, 1995.

[6] R. Hartley A. Zisserman. *Multiple View Geometry*. Cambridge University Press, Cambridge, 2000.

[7] M. Isard and A. Blake. Visual tracking by stochastic propagation of conditional density. *Proc. 4th European Conf. Computer Vision*, pages 343–356, April 1996.

[8] M. Isard and A. Blake. A smoothing filter for condensation. *Proc 5th European Conf. Computer Vision*, 1:767–781, 1998.

[9] J. P. Luck, C. Debrunner, W. Hoff, Q. He and D. E. Small. Development and analysis of a real-time human motion tracking system. *WACV*, pages 196-202, 2002.

[10] J. P. Luck, W. Hoff, D. Small and C. Little. Real-Time Markerless Human Motion Tracking using Linked Kinematic Chains. *JCIS*, pages 849-854, 2002.

[11] J. MacCormick and A. Blake. Partitioned sampling, articulated objects and interface quality hand tracking. *Accepted to ECCV*, 2000.

[12] J. MacCormick. *Stochastic Algorithms for Visual Tracking*. Springer-Verlag, London, 2002.

[13] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human Body Model Acquisition and Tracking Using Voxel Data. *IJCV*, 53(3):199-233, 2003.

[14] C. Sminchiescu and B. Triggs. Estimating Articulated Human MotionWith Covariance Scaled Sampling. *International Journal of Robotics Research*, 22(6):371-393, 2003.

[15] D. E. Small. *Real Time Shape from Silhouette*. Masters Thesis. University of Maryland, 2001.

[16] A. W. B. Smith and B. C. Lovell. Autonomous Sports Training from Visual Cues. *ANZIIS*, 2003.

[17] "Dragon Fly at Point Grey Research". [Online] Available at http://www.ptgrey.com/products/dragonfly/dragonfly.pdf, last accessed 25/8/2004.