Autonomous Sports Training from Visual Cues

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

Abstract

Computer driven biometric analysis of athlete's movements have proven themselves as effective sports training tools. Most current systems rely on the use of retro-reflective markers or magnetic sensors to capture the motion of the athlete, so the biometric analysis can be performed. Video based training tools have also proved to be valuable instructional aids, however most require significant human interaction for analysis to be performed. This paper outlines an ongoing project focussed on capturing posture without the use of any markers or sensors, while still capturing enough information for an automated analysis to be performed. The approach taken to solving this problem is presented, as well as the current state of development of a an instructional aid for golfers.

1. Introduction

Biometric analysis has established itself as an effective training tool for athletes. Currently 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. An example of this is the 3D-GOLF™system[11], where posture is obtained from magnetic sensors and then professionally analyzed. There is an inherently large setup time involved with the use of these systems, increasing the cost and inconvenience to the athlete. Were the athletes postural information captured visually, this setup time could be greatly minimised. Furthermore, vision based systems have the potential to be produced at much lower cost than magnetic sensor systems due to the mass production of cheap cameras sensors. A final advantage of obtaining posture from visual cues is that recorded video of the actual athlete can be shown to provide feedback, instead of a 'stickman' representation or an animated model.

Video based analysis tool have been widely used as instructional aids. Systems such as Swingcam[14] and the GolfTek®Video Swing Analysis System[12] record footage of a golf swing for the player to view. Analysis is left to be done by the player in these systems. Other systems can perform limited analysis, however many require human processing to determine the position of important features. Motion Coach™[13] is one example of this.

With continuous developments in the image processing field and increases in the computational power of PC's, visual motion capture and analysis applications have become quite feasible. An example of such a system is the virtual personal aerobics trainer[4], which provides feedback on exercise activities from purely visual cues.

The aim of the project is the development of a system which uses visual cues to obtain an athlete's postural information, and analyzes this with respect to a learned ideal motion. Completely automated feedback can then be given based on differences between the athlete's motions and the technically correct motion. No human interaction is required in this process, giving it an advantage over other systems. Golf has been chosen as the sport of focus as there is a well defined technically correct swing, and because of the golfer's limited motion in space (a camera can see the whole swing without moving). The visual system can also be integrated with golf swing analyzer mats, which provide data that the visual system cannot detect.

2. System Overview

The system envisaged is one where the golfer addresses a ball situated on a swing analyzer mat. The swing analyzer mat is a commercially available product that has been used as a golf instruction aid for over 20 years, and is discussed in section 2.1. The golfer will be facing a single camera (current focus is on a purely 2D system) with a display monitor mounted above. This system is illustrated in figure 1. Once the golfer has completed their swing, information collected from both the camera and the golf mat will be combined to provide both audio and visual feedback on the monitor for the golfer to observe. The project can be though of as a fusion of the measurements from the analyzer mat and the visual system. Analyzer mats have been used to determine what happens around impact time, but not why this has occurred. We are adding analysis of the postural variation which caused the given swing. It is envisaged that training booths equipped with this system could be incorporated into driving ranges as instructional aids.

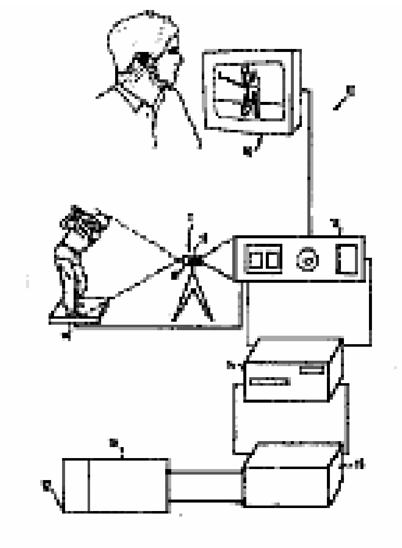


Figure 1. Original concept of Golf Swing Personal Trainer from Hi-Tech Video's 1991 Patent Application [15]

The camera used in this system is the dragonfly firewire camera from Point Grey Research[10]. This camera is capable of 30 frames per second, at a resolution of 640x480 pixels.

2.1. The Swing Analyzer Mat

Golf swing analyzer mats use an array of sensors to detect the club passing overhead. Top of the line units use around 53 infrared sensors which measure:

1. club head path
2. club face angle
3. club speed
4. impact location
5. tempo
6. ball carry
7. ball trajectory (push, pull, slice, straight, draw etc)

and then determine the result of the shot — either fairway or rough.

The golf analyzer mat adds extra capabilities for feedback as it can detect small changes in the club head angle which are infeasible to detect via visual means. For example, if the visual analysis detects no significant deviations from the correct swing, yet the predicted path of the ball is not straight, the mistake can be asserted to be caused by an incorrect rotation of the wrist. Imagine this rotation causes the club to be rotated about its axis, something the visual analysis will not detect but will result in a poor shot. The program can infer this and provide feedback showing the frame where the club impacts the ball, appropriately zoomed, allowing the player to perform their own comparison between their wrist and the professional's.

We have modified a GolfTek®Pro III Golf Swing Analyzer for use with our system. Details on this analyzer are given at [12]. Data is sent to an Atmel AT90S8515, which interfaces with a PC using the RS232 protocol. Consideration is being given to making this a wireless link.

2.2. Off site processing

It is envisaged that analysis performed using this system will be performed at a central site. Sensor data will be transmitted to a central site via technology such as broadband communications, the analysis performed, and the result transmitted back to the user.

2.3. Intended Application

It is intended that this system would be used by social, rather than professional, players. As such it is aimed at players looking to improve their game — this would occur by either supplementing coaching from a golf professional, or developing a level of prowess that allows them to take full advantage of the professional's advice. It is not intended that this system should provide the in-depth analysis required by professionals.

2.4. Adding expert players to the learned set

While the initial design will focus on providing feedback based on a technically correct swing, it is intended that expert golf professionals' swings will be added to this database. Professional golfers swings are not always what is considered technically correct. An example of this is Jim

Furyk's swing, which is considerably different from a technically correct swing. Despite not being technically correct, Furyk's swing was good enough to win the 2003 U.S. Open. A player could try to replicate his swing, or another of their favorite players.

3. Visual Tracking Overview

In this application we need an algorithm capable of robustly tracking the human body through the range of motion encountered during a golf swing. It should use minimal knowledge of the background as we expect the background to change as people move behind the golfer.

For the algorithm to track robustly, we assert that it should search a wide range of the search space, as opposed to simply finding the nearest local maxima in the search space. This will enable tracking through cluttered environments. Algorithms such as AAMs[3] or Kalman filtering [9] do not satisfy this conditions, as they both seek out the nearest local maxima in the search space to the expected position. This can just as easily be an artifact of clutter as the true object. Isard and Blake's condensation[5] algorithm was designed to meet just these criteria, and not rely on assumptions about the background, though they can prove useful.

Condensation is a stochastic algorithm which uses a particle filtering approach to determine the most probable target configuration. Each particle is a spline based contour constrained to lie in a configuration space. The sampling nature of the particle filter attempts to search the entire configuration space, not just a local region of it. Considering the case where the number or particles, $n = \infty$. In this case the entire configuration space will be searched. Although the amount of samples must be restricted to a practical number, condensation provides the ability to search a wide area of the configuration space, and not just find the nearest local maxima. This gives it the ability to track in applications where the posterior distribution is multi-modal, i.e., any cluttered environment.

3.1. An Overview of the condensation algorithm

The steps of the condensation algorithm from [5] are as follows for each time step t:

1. Starting with a set of particles $\mathbf{s}^{(t-1)} = (s_1^{(t-1)}, ..., s_n^{(t-1)})$ with prior probabilities $\boldsymbol{\pi}^{(t-1)} = (\pi_1^{(t-1)}, ..., \pi_n^{(t-1)})$, a new set of samples are randomly drawn via a weighted sampling operation.

2. Dynamics are then applied to each particle in the set, giving the new sample set $\mathbf{s}^t$.

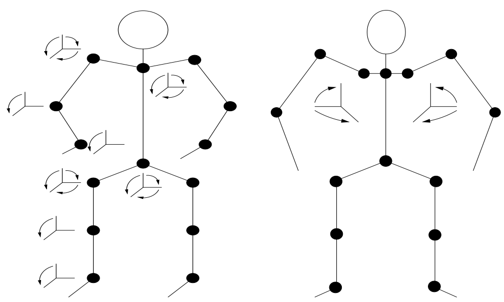
3. The posterior $\boldsymbol{\pi}^t$ for the sample set $\mathbf{s}^t$ is constructed by finding the observational likelihood of each particle

4. Estimates of the object position are calculated.

The dynamics used in this project are currently a learnt 2nd order auto-regressive process, as described in [2]. Observational likelihoods are constructed by drawing an evenly spaced set of measurement lines along curve normals of each contour (remember each particle is a contour). MacCormick [7] gives a number of methods to calculate probabilities based on the set of features on each measurement line of a contour. Currently the Poisson method with robust color detection is being used. The spline position can be calculated using MAP, MLE or MMSE estimates. As in [5] we are using a MAP estimate of the target position, obtained by an average of all the particles weighted by their probabilities.

3.2. Modelling the golfer

For the initial development of this application a principal component approach has been used to determine the configuration space. This is due to uncertainty in the nature of the articulated model we wish to use. An example to show why this uncertainty exists focuses on the players shoulders. Take for example, a 26 DOF (degrees of freedom) articulated model. This model would assume the shoulder carapaces are in a fixed position relative to the top of the spine. During a movement like shrugging however, the shoulder carapaces violates this assumption. Consider trying to swing a golf club with your shoulders in a raised position like when shrugging. Clearly a posture like this will lead to a poor swing, necessitating the need detect such postural variations. Another two degrees of freedom need to be added to model this. Figure 2(a) shows a 26 DOF model, while figure 2(b) shows the incorporation of 4 DOF to model the movement of the shoulder carapace, which models both the shrugging motion and the hunching of the shoulders.



(a) 26 DOF model (b) 30 DOF model

Figure 2. Modelling the articulated body

Once data from a training set has been analyzed, the articulated model which best explains the variation can be chosen.

With the articulated model chosen, a surface must be fitted to each link. This presents another reason why we are starting with the principal component approach as the surface must represent the clothes and not just the skin. This causes variability in the shape of each surface during the swing, to be investigated at a later stage.

3.3. Coping with the high dimensionality configuration space of the human body

One problem associated with tracking the human body is the high dimensionality of the configuration space. The higher dimensionality of the configuration space, the more particles that are required, and hence the higher the computational cost. MacCormick and Blake [8] 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. Clearly a large number of dimensions (around 26 for an articulated human) creates the need for a impractical number of samples. To make tracking in high dimensional spaces tractable, methods such as annealed particle filtering [1] have been proposed. This method allows tracking in high dimensional spaces assuming no knowledge of motion, angle of view, or availability of labelling cues. The condensation algorithm propagates the conditional density $p(X|Z)$ using $p(Z|X)$, where Z is the image and X the configuration space, whereas annealing finds the maximum value of a weighting function $w(Z, X)$. The benefit of this is that $w(Z, X)$ is computational cheaper and requires fewer evaluations, however this comes at the cost of not being able to work within a robust Bayesian framework. In the case of the golf tracking application these assumptions can be relaxed, the angle of view is known and importantly we can assume that the head is never occluded during the swing from this angle. This allows a hierarchical search to be performed using the knowledge that:

1. the head will not be occluded
2. the only object which occludes the arms and club is the head
3. the only objects which occlude the legs are the arms and club

MacCormick and Isard [8] develop a method of partitioned sampling, whereby the configuration space is decomposed into multiple partitions to be search in a hierarchical order. Dynamics are applied to the first partition, the result being multiplied by the observational density of this partition. A resampling operation is then performed, and the dynamics of the second partition applied. Now we multiply by the observational density of the second partition. The process is repeated until all partitions have been calculated. It was shown in [8] to reduce the number of particles required to

$$N \geq \frac{D_{min}}{\alpha}. \tag{2}$$

Another advantage of this method is that the number of particles used in each partition can be varied.

Considering two partitions in this application, we have one which contains the head, and another which contains the arms. The partitions meet at the tops of the shoulders. These are drawn as individual contours in figure 3. The observational density of the head partition is independent of the arms partition, and hence can be found with no knowledge of the location of the arms. The prior can the be resampled weighted by this observational density. This concentrates the search in areas of the arms partition which correspond to high observational densities in the head partition.

(a) The head partition

(b) The arm partition

Figure 3. Different partitions of the configuration space, manually annotated

3.4. Smoothing the output

The nature of this application allows for post processing of the tracking results, as the analysis occurs after the golfer has completed their swing. Because of this, a backwards optimization phase can be added to the algorithm to improve results. Isard and Blake [6] present a framework for a output smoothing filter. From section 3.1 the output of a time step of the condensation algorithm consists of a set of state vectors $\mathbf{s}^t = (s_1^t, ..., s_n^t)$ with associated probabilities $\boldsymbol{\pi}^t = (\pi_1^t, ..., \pi_n^t)$ at time t. The smoothing algorithm can be thought of as finding the best path through a Hidden Markov Model, where the transitional probabilities are derived form the dynamic model.

This smoothing framework provides a powerful tool when multiple hypotheses of the object position are present. In this case, the prior for the time step will be multi-modal. Should the posterior reflect the same peaks as the prior, we conjecture that a MAP estimate of the object position will be not be useful. The MAP estimate will average the competing hypotheses, producing a result in the middle of them that does not necessarily represent a probable object position. The tradeoff however is the restriction upon the type of dynamic model which can be used, i.e it must be first order to fulfill the Markov criteria. A less accurate dynamic model will increase the number of particles required for successful tracking. At the same time it is worth remembering that as this is a training aid, deviation from the dynamic model is exactly what we are looking for. Although strong multiple hypotheses are not expected in this application, it will be determined at a later stage if this smoothing technique proves beneficial.

4. Providing feedback to the golfer

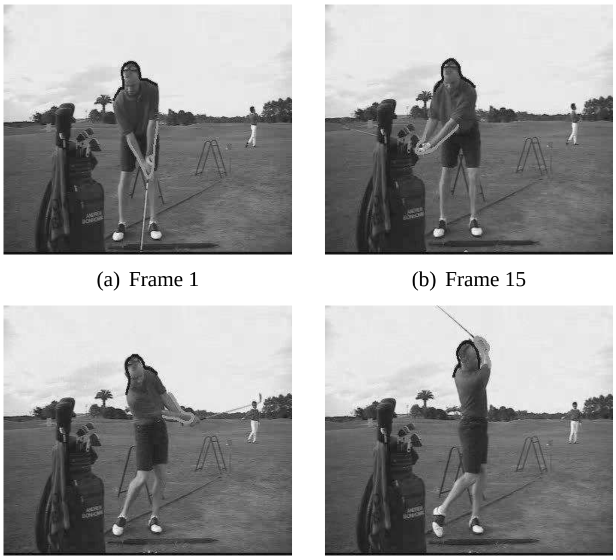
Once the results from the visual tracking have been obtained, feedback can be provided to the golfer. To do this, the system will have a learned set of swings from technically correct professional golfers. Areas where the players swing varies significantly from the mean of the learnt set can then be highlighted. The correct postural position can be superimposed onto the video of their swing, or shown alongside the video. Audio advice will also be provided, again from a learnt set, on causes of the problem and the expected result of the shot as determined by the analyzer mat.

4.1. Matching the timing

To match the timing of the players swing and the learned set, the swing is broken into two parts. These are the back swing and the down swing. By detecting when the player is at the top of their back swing, indicated by a change in direction of arms, the learned set data can be interpolated or extrapolated to match the time taken for the player to reach that point. A rule of thumb is that the down swing should be twice as fast as the back swing and is referred to as the tempo of the swing. The down swing from the learnt set can also be interpolated or extrapolated to match the player's swing speed, or the player can try to match the learnt set to improve their tempo.

5. Current state of the project

Contours outlining the head and arms have been successfully tracked throughout the golf swing. These result are shown in figure 4. Tracking here was achieved using the condensation algorithm, with the Poisson observational likelihood function function described in [7]. The hierarchical tracking algorithm described in section 3.3 was not used in this test, meaning each contour was tracked independently of the other.



(a) Frame 1

(b) Frame 15

(c) Frame 30

(d) Frame 38

Figure 4. Results of tracking

6. Future Work

The focus of the project now is on creating a concept demonstration version of the product to arouse interest in the golfing community. To achieve this a small set of swings taken from available golf players will be learnt, allowing for rudimentary feedback to be given. This feedback is envisioned to be simple suggestions such as advising a player not to lift their head or to keep their arms straight. It is hoped that this concept demonstrator version can be completed around January 2004.

References

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

[2] A. Blake, and M. Isard. *Active Contours*. Springer-Verlag, London, 1998.

[3] T.F. Cootes, G.V. Wheeler, K.N. Walker and C. J. Taylor. View-based active appearance models. *Image and Vision Computing*, (20).

[4] J. Davis and A. Bobick. Virtual pat: a virtual personal aerobics trainer. *MIT Media Lab Perceptual Computing Group Technical Report*, (436), 1998.

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

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

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

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

[9] R. Szeliski and D. Terzopoulos. Tracking with kalman snakes. *Active Vision*, pages 3–20, 1992.

[10] "Dragon Fly at Point Grey Research". [Online] Available at http://www.ptgrey.com/products/dragonfly/dragonfly.pdf, last accessed 25/8/2003.

[11] "Golf Bio Dynamics". [Online] Available at http://www.golfbiodynamics.com/html/GAS.htm, last accessed 20/8/2003.

[12] "GolfTek". [Online] Available at http://www.golftek.com, last accessed 20/8/2003.

[13] "Motion Coach". [Online] Available at http://motioncoach.com/mp_rmvbck.htm, last accessed 22/8/2003.

[14] "Swing Cam". [Online] Available at http://www.swingcam.co.uk/index.htm, last accessed 20/8/2003.

[15] International Patent Classification(s)A63B 069/36 G06F 015/62, Application No.:WO92/21412, Application Date:01.06.92, PCT Publication Number:WO92/21412, Priority Data Number: PK6405, Date: 30.05.91, Country: AU AUSTRALIA, Publication Date: 08.01.93, Applicant: Richard John Baker, Inventor(s): Richard John Baker.