Visual Golf Club Tracking for Enhanced Swing Analysis

Nicolas Gehrig Vincent Lepetit Pascal Fua

Computer Vision Laboratory (CVLab)
Swiss Federal Institute of Technology (EPFL)
1015 Lausanne, Switzerland
{nicolas.gehrig, vincent.lepetit, pascal.fua}@epfl.ch

Abstract

This paper presents a new visual tracking technology that relies on the use of a global motion model to achieve robustness. We demonstrate its effectiveness for the purpose of retrieving the 2D spatio-temporal trajectory of a golf club head from ordinary video sequences of golf swings, so that information about club orientation, local speed and acceleration can also be obtained. We have integrated it into a fully automated system that requires neither user intervention nor the use of instrumented golf club or clothing, and that is usable in a natural environment with a potentially cluttered background. Our algorithm robustly fits a global swing trajectory model to club location hypotheses obtained from single frames. This process makes our approach very robust and it will soon be integrated into a commercial product. Several experimental results are presented to illustrate the success of this new method.

1 Introduction

The use of video in sports training sessions is considered to be a very useful tool by many coaches and athletes. The opportunity for athletes to watch their own performance on screen can help them discover and better understand their strengths and weaknesses. However, the interaction with the video sequence is usually fairly limited, consisting mainly of slow motion replays. Therefore, there is great interest in enhanced analysis tools that provide more quantitative information and more interactivity.

This paper presents a robust visual and fully automated tracking technique for retrieving the 2D spatio-temporal trajectory of a club head during a golf swing from "face on" video sequences such as the ones shown in Figure 1. Since a swing is usually very fast, it is also hard to track, especially against a potentially cluttered background. Our contribution is therefore an algorithm that efficiently combines a robust approach to club detection with a global motion model to achieve both automation and reliability in a natural environment. Figure 1 depicts a subset of the database of 35 swings against which we have tested it.

Our system thus provides the golfer with visual information that can be used to analyze and compare swings. Useful information such as local speed and acceleration along

BMVC 2003 doi:10.5244/C.17.47

Figure 1: A selection of some results of our tracking system. Color-coded trajectories showing the local speed of the club head in terms of different speed ranges.

the swing trajectory can also be gathered, allowing for very precise comparison of different swings, not only in terms of spatial trajectory but also in terms of temporal evolution.

2 Related Work and Approach

Many visual tracking techniques have been proposed in the literature since the beginning of Computer Vision. The usual approach is recursive: the target position and shape in the current frame are first predicted from its estimated state in the previous frame, and

then adjusted based on image observations. Many probabilistic approaches using particle sets such as the *Condensation* [1] algorithm are also very popular for dealing with more complex tracking. *Data Association* [2] approaches are extensively discussed in discrete targets tracking literature, and seem to be more suitable for our problem of tracking golf clubs on cluttered background than *Condensation*.

Using these techniques in a practical setting remains, however, quite difficult, and very few of those are available as commercial products. The main problem is that they tend to suffer from a lack of robustness. All of these approaches consider recursive motion models where the current state of the target can be estimated from its previous state as: $X_{t+1} = f(X_t)$. Their behavior is therefore very local, and it is thus very difficult to consider a global motion model such as defined by a golf swing. Our method addresses this problem by introducing a global motion model.

We first process the whole video sequence to create plausible hypotheses for the position of the club head in each frame. This is a difficult task because the club head is usually very small and has no well-defined color or shape. Hence, instead of directly looking for it, we extract the straight part of the club, or shaft. Since the head remains at the shaft's extremity, it becomes easier to find it. The shaft is detected by looking for a moving thin and straight object. The extraction is thus based on a motion detection followed by a parallel straight edges detection. This provides us with a complete set of possible shaft positions in each frame of the sequence. Although the club is often successfully extracted, it is usually not the only thin object detected in the scene, resulting in several false-alarms in many frames.

We then process the whole set of hypotheses in order to locate important events in the sequence such as the beginning and the end of the swing as well as the transition between upswing and downswing. As shown in Figure 2, upswing refers to the motion from somewhere close to the ground to the top position, and downswing refers to the motion from the top position back down.

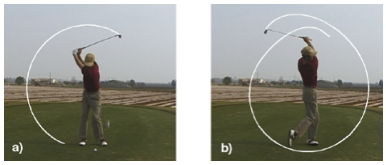

Figure 2: The golf swing is decomposed into (a) upswing and (b) downswing.

Finally, upswing and downswing trajectories are obtained by robustly fitting a polynomial curve to club detection using a RANSAC [3] like algorithm. We now turn to individual components of our approach.

3 Club Extraction

Detecting and extracting a specific object in an image is usually not an easy task. Color and shape features are most often used to perform the detection. In our case, a golf club does not have any specific color and is highly reflective. Furthermore, because it is so thin it may be blurred when its velocity is high.

To detect it, a motion detector is first applied to the current frame. For that purpose, the difference between the current and the previous frames is computed as the euclidean

distance in the *YUV* color-space and the result is thresholded, producing a binary mask representing the moving objects between these two frames. A morphological closing operation is also applied in order to fill small gaps in the extracted motion regions and to smooth their borders. The same operation is also applied to the current and next frames producing a second binary mask. Finally, a logical bitwise AND operation between these two masks gives the mask of the moving objects in the current frame. Thus,

$$M_t = C_2(H_T(I_t - I_{t-1})) \cap C_2(H_T(I_t - I_{t+1}))$$

where M_t is the final binary mask of the motion regions at time t, I_t is the current image at time t, H_T is a thresholding operation with threshold T, and C_n consists of n successive morphological dilatation and erosion operations.

Next, *Canny* edge detection [4] is applied on the moving regions of the image from which the method tries to extract straight components. For that purpose, chains of adjacent pixels are first extracted from the detected edges. Then, straight segments are obtained from an exhaustive search along these chains using a fixed tolerance for their straightnesses. The shaft usually produces a pair of close parallel segments. So, each such pair of detected segments is then selected and merged into a single one. Unfortunately, only part of the shaft is usually retrieved, for example because it is slightly bent or because its extremity blurred enough to be almost undistinguishable from the background. Post-processing needs therefore to be applied to each detected segment in order to recover the position of the club head and the golfer's hands with accuracy. Since it is not known at this time which extremity of the segment corresponds to the club head, we clone all segments and assign them opposite directions from the originals. Figure 3 illustrates this extraction process.

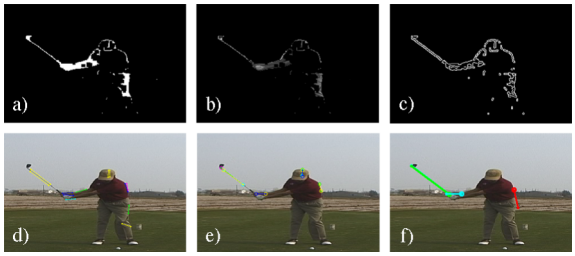

Figure 3: Hypotheses generation. (a) Binary mask obtained from the motion detection. (b) Current image converted into grayscale and covered by the mask. (c) Result of the *Canny* edge detection. (d) Detected segments. (e) Each pair of close parallel segments are merged into a single segment. Then, each resulting segment is processed, trying to retrieve the club head and the hands positions with accuracy. (f) Remaining hypotheses after the rejection tests.

Although this detection presents a good rate of success in extracting the current club position, false-alarms are inevitable. To limit their number, we analyze all hypotheses and get rid of all of them presenting a physically impossible position. For example, too short or too long detected clubs can easily be removed. All remaining hypotheses are stored and will be processed in the next part of the method to estimate the swing trajectory.

4 Trajectory Estimation

This is the heart of our approach and is key to achieving robustness. It is designed to retrieve the swing trajectory from the set of hypotheses which can contain many outliers. The first step consists in analyzing the hypotheses in order to localize the upswing and downswing regions in the sequence. Then, the trajectory estimation is performed independently on both of these regions.

4.1 Trajectory Model

Before introducing our estimation procedure, we need to formalize the choice of our swing trajectory model. Our goal is to find a simple model able to represent any club head trajectory from any golf swing with the smallest possible number of parameters that yields a sufficient level of precision.

We assume that all the golf swings that we want to track are made up of an upswing and a downswing. We propose two models to represent these trajectories. Two functions $\rho_{up}(\beta)$ and $\rho_{down}(\beta)$ are defined in polar coordinates using a central point as origin. The exact location of this reference point has no real influence on the final results. However, it should be placed roughly in the center of the trajectory. These two functions give the distance ρ between the club head and the reference point as a function of the angle β. This angular value β is defined with a vertical origin (looking to the top of the image), increasing clockwise. The upswing trajectories are defined on the range $\beta \in [\pi, \frac{7\pi}{3}]$, while the downswing trajectories are defined on the range $\beta \in [\frac{7\pi}{3}, -\frac{\pi}{2}]$. Figure 4 presents these trajectories and the defined referential.

In some cases, β may need to be adjusted by a value of $\pm 2\pi$ in order to be coherent with the current region of the swing. We want that all hypotheses belonging to the upswing and the first part of the downswing are in the range $\beta \in [\pi, \frac{7\pi}{3}]$. We also want that all hypotheses belonging to the last part of the downswing are in the range $\beta \in [\pi, -\frac{\pi}{2}]$. These conditions ensure a continuous evolution of β during the whole swing.

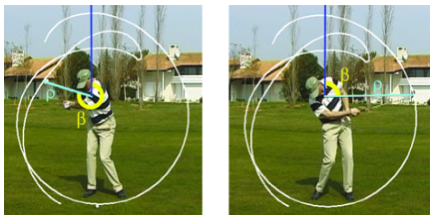

Figure 4: Polar coordinates of the club head using our defined referential.

We manually acquired many different golf swing trajectories in order to analyze the behavior of these two functions. We observed that they can usually be very easily and precisely approximated by simple polynomial functions of rather small degrees. The idea of our trajectory estimator will thus be to find such a polynomial function of a certain fixed degree matching with a hypothesis in the highest number of frames.

An important question is how can we determine the optimal degrees to use for these two polynomial functions. Too small degrees will not allow for a precise representation of all possible club head trajectories, while too large degrees may present unstable behavior during the estimation procedure in the presence of outliers, due to a too large number

of degrees of freedom. There is therefore an important tradeoff in the choice of these degrees. A reasonable choice would thus be to determine the smallest degrees providing acceptable precision for any swing trajectory approximation.

We tried to estimate an important number of manually acquired swing trajectories with polynomials of different degrees. For each estimation, we computed its mean square error and support values. The support is computed as the percentage of acquired points closer to the estimated curve than a given threshold T. Averaging our results on many different swings, we observed that the *MSE* drops to a relatively small value for an upswing estimation of degree 4 and a downswing estimation of degree 6 (see Figure 5). Moreover, the support reaches quite important values for the same degrees. These two degrees seem therefore to be optimal choices in order to deal efficiently with the present tradeoff.

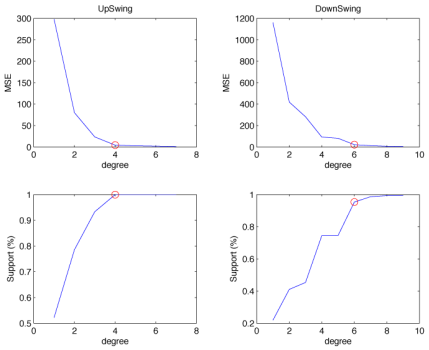

Figure 5: Determination of the best degrees to use for the polynomial trajectories estimations. (Averaged results on an important set of different trajectories) Degrees 4 for upswing and 6 for downswing are obtained from this observation. These two values are the smallest ones allowing for a relatively precise representation of any swing trajectory.

4.2 Temporal Segmentation of the Sequence

We want to localize the upswing and downswing regions in the sequence. For that purpose, we need to analyze the set of hypotheses and identify several key events such as the position of the beginning of the swing, the limit between the upswing and the downswing and the end of the swing. Analyzing the average elevation (the y coordinate) of the club head's hypothetical position along the sequence, we can get an idea about the evolution of the altitude of the real club head. Assuming a stationary distribution of the position of the outliers, it is thus possible to statistically estimate the evolution of the positions of the inliers. We create a sequence $s[n]$ containing the average elevation of the hypotheses in each frame and we filter it several times with a simple average filter $f[n]$. $s[n]$ is defined such that $s[i] = \frac{1}{N_i}\sum_{j=1}^{N_i} h_{i,j}$ for $i \in [I_{firstFrame}, I_{lastFrame}]$, where N_i is the number of hypotheses in the ith frame and $h_{i,j}$ is the altitude (y coordinate) of the jth hypothesis of the ith frame. Whenever $N_i = 0$, we set $s[i]$ to a certain fixed constant. The filter is defined such that $f[n] = \frac{1}{M}$ for $n = 0, \ldots, M-1$.

The resulting sequence usually presents a nice and smooth curve corresponding quite well to the club head altitude evolution, as shown in Figure 6. We can therefore retrieve the desired bounds corresponding to some easily identifiable peaks. A simple procedure looks for the top position of the club corresponding to the first peak higher than a certain threshold. Then, it searches backwards for the start position and forwards for the end position. It also estimates the position corresponding to the time when the club hits the ball.

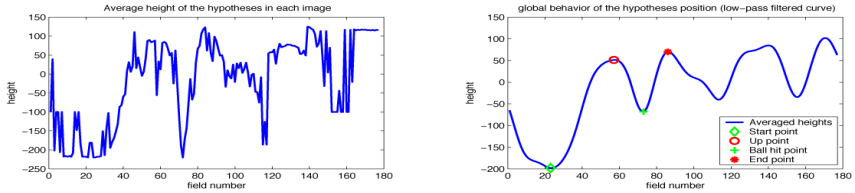

Figure 6: Time bounds retrieval process. The first graph shows the average elevation of the hypotheses for each frame. The second graph is the result of this sequence filtered three times with $f[n]$. A simple analysis of this result allows for a precise estimation of the desired time bounds.

Using these results, we can now apply a more restrictive rejection test to the remaining hypotheses. Depending on the current region, we can get rid of some remaining outliers that do not represent a physically valid position for that region. Decreasing the number of outliers will obviously help the trajectory estimation, therefore guaranteeing a very high success rate.

4.3 Robust Trajectory Estimator

Assume that we want to find a polynomial upswing trajectory estimation $\hat{\rho}_{up}(\beta)$ of degree d_{up} and a downswing trajectory estimation $\hat{\rho}_{down}(\beta)$ of degree d_{down} (We typically use $d_{up} = 4$ and $d_{down} = 6$ as defined in Section 4.1). The algorithm (*RANSAC*-like) proceeds as follows. It randomly chooses one hypothesis in $N_{up} = d_{up} + 2$ distinct frames belonging to the upswing region, such that the β value of the hypotheses strictly increases in the range $[\pi, \frac{7\pi}{3}]$ when looking at them in chronological order. Then, it determines the best polynomial function of degree d_{up} fitting these hypotheses in the mean square error sense.

Let (β_i, ρ_i) be the polar coordinates of the club head of the ith randomly selected hypothesis. We want to find the coefficients $\underline{c} = [c_0, \ldots, c_{d_{up}}]^T$ of a polynomial function such that $\|A\underline{c} - \underline{\rho}\|^2$ is minimal, where

$$\mathbf{A} = \begin{bmatrix} 1 & \beta_0 & \ldots & \beta_0^{d_{up}} \\ 1 & \beta_1 & \ldots & \beta_1^{d_{up}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \beta_{N_{up}} & \ldots & \beta_{N_{up}}^{d_{up}} \end{bmatrix} \quad \text{and} \quad \underline{\rho} = \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_{N_{up}} \end{bmatrix}$$

Using the Pseudo-Inverse theorem (and assuming A of maximum rank), we know that this minimum is obtained for $\underline{c} = [A^T A]^{-1} A^T \underline{\rho}$.

Once we have this estimation at our disposal, we check for each frame of the upswing region if it contains a hypothesis close to this trajectory (closer than a certain threshold T_{up}). We compute the support S of this estimation as the number of frames in which such a close hypothesis is present and we compute a distance value as the mean square distance of these close hypotheses to the estimated trajectory.

This whole process is repeated many times and the estimation presenting the highest support S is kept. By finding an important sequence of hypotheses corresponding to a

trajectory defined by a smooth polynomial function, we can ensure that these hypotheses are inliers.

The final upswing trajectory is redefined using the S hypotheses belonging to the support of the best estimation found. We compute the polynomial function best fitting all these hypotheses in the mean square error sense as previously. The coefficients $\underline{c}$ of the polynomial function $\hat{\rho}_{up}(\beta)$ are thus obtained as $\underline{c} = [A_{big}^T A_{big}]^{-1} A_{big}^T \underline{\rho}_{big}$, where

$$\mathbf{A_{big}} = \begin{bmatrix} 1 & \beta_0 & \dots & \beta_0^{d_{up}} \\ 1 & \beta_1 & \dots & \beta_1^{d_{up}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \beta_S & \dots & \beta_S^{d_{up}} \end{bmatrix} \quad \text{and} \quad \underline{\rho}_{\mathbf{big}} = \begin{bmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_S \end{bmatrix}$$

Once the final upswing trajectory estimation $\hat{\rho}_{up}(\beta)$ is defined, we use it to adjust the previously estimated temporal position between the upswing and downswing. Then, the same approach is used to find the estimation $\hat{\rho}_{down}(\beta)$ of the downswing trajectory.

4.4 Speed Estimation

In the previous section, we proposed a method to robustly retrieve the spatial trajectory of the club head. Here, we present an additional approach used to estimate the temporal evolution of the club head along this trajectory. Let the two functions $\beta_{up}(t)$ and $\beta_{down}(t)$ correspond to the temporal evolution of the angular coordinate β of the club head during the upswing and downswing. We want to approximate these functions with two polynomial functions $\hat{\beta}_{up}(t)$ of degree d_{up}, growing in the range $[\pi, \frac{7\pi}{3}]$ and defined for $t \in [t_{start}, t_{up}]$, and $\hat{\beta}_{down}(t)$ of degree d_{down}, decreasing in the range $[\frac{7\pi}{3}, -\frac{\pi}{2}]$ and defined for $t \in [t_{up}, t_{end}]$. The best polynomial degrees to use for these temporal estimations have been determined using the same approach as for the trajectory estimations (see Section 4.1), and are $d_{up} = 3$ and $d_{down} = 5$. The time indexes t_{start}, t_{up} and t_{end} correspond respectively to the beginning of the swing, the limit between upswing and downswing and the end of the swing.

In order to find a good approximation of $\beta_{up}(t)$, we use all the S hypotheses belonging to the support of our estimated trajectory $\hat{\rho}_{up}(\beta)$. Let β_i be the β coordinate of the ith selected hypothesis and t_i be the index of the frame containing it. We find the coefficients $\underline{c} = [c_0, \dots, c_{d_{up}}]^T$ of the best polynomial estimation $\hat{\beta}_{up}(t)$ as $\underline{c} = [D^T D]^{-1} D^T \underline{\beta}$, where

$$\mathbf{D} = \begin{bmatrix} 1 & t_0 & \dots & t_0^{d_{up}} \\ 1 & t_1 & \dots & t_1^{d_{up}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_S & \dots & t_S^{d_{up}} \end{bmatrix} \quad \text{and} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_S \end{bmatrix}$$

The same approach is then used to compute the estimation $\hat{\beta}_{down}(t)$.

Having $\hat{\rho}_{up}(\beta)$, $\hat{\beta}_{up}(t)$, $\hat{\rho}_{down}(\beta)$, $\hat{\beta}_{down}(t)$ and the three time indexes t_{start}, t_{up} and t_{end}, we have a complete spatio-temporal model of the whole swing trajectory. This model can be used to draw the swing trajectory when playing the sequence or to compare different trajectories. Local speed and acceleration can also easily be derived from it.

5 Experimental Results

As shown in Figure 1, the system has been tested on many different sequences from various players in various environments. The video sequences were interlaced DV-PAL encoded. The PAL system has a frame rate of 25 frames per second, which means that 50 fields are captured every second. Such a high acquisition frequency is quite important in order to provide sufficiently close detections for precise trajectory estimation. Nevertheless, in some sequences, the club head reaches a speed of about 160km/h just before hitting the ball. At such a speed, the displacement of the club head is of about 90cm between two consecutive fields, which is quite large, meaning that only few club head positions are available during this part of the swing.

Our system behaved successfully on most of the test sequences, including those presenting a very fast swing. The sequences on which the system failed to extract a correct trajectory presented some foreseeable problems. For example, the shutter speed used for recording most of these sequences was too slow, producing extremely blurred clubs during the fast parts of the swings, making it impossible for the club extraction to retrieve them. Too few detections were therefore available to estimate the correct trajectory. Our system is very robust, given a reasonable quality for the video sequences. Actually, whenever the club is relatively well detected throughout the whole swing (no long sequence of misdetections), the trajectory is correctly returned, even in the presence of a high number of outliers. Figure 7 presents a wrong trajectory estimated from a sequence acquired with a too slow shutter speed.

Figure 7: Trajectory estimation failure due to video sequence quality problem. (a) The estimated trajectory is seriously wrong during the fast part of the downswing. (b) Club position at the beginning of the upswing; since it is moving slowly, the club is clearly represented on the video sequence and will be correctly extracted. (c) Club position just before the ball hit; due to its high velocity and the too slow shutter speed of the acquisition system, it is almost undistinguishable from the background and won't be detected during the club extraction. Having no correct detection in this area, it will therefore be impossible for the trajectory estimation to find the correct trajectory.

The estimated trajectories present a good level of precision. No deflection between them and the real club head positions can usually be noticed. Only during the fastest part of the swings have some precision problems been observed. Since only few frames are usually available in these regions, an abrupt change of direction or acceleration is hardly detectable. Even they, the estimated trajectory is usually off by no more than ten pixels, which is quite reasonable. Obtaining the speed and acceleration of the club head just before it hits the ball with a higher precision might be very interesting, but would require the use of a video camera with a higher acquisition frequency.

6 Conclusion

We have proposed a new approach to tracking that is applicable when a global trajectory model is available. This approach relies on robustly fitting a polynomial trajectory model to a set of detections. It makes the tracker robust and automated enough to be integrated into a commercial product that does not require particular knowledge from the user and that is designed to work in true outdoor environments.

We believe that our approach is very general and can be extended to more complex motion models using splines or PCA-based [8] representations. It can therefore be a valuable alternative to standard tracking approaches for any application where a specific motion has to be tracked, which is the case for many sports gestures, and more generally in all training situations in which one deals with specific gestures that are parts of known procedures.

In future work, we plan to acquire swings from multiple golfers of different skills, and classify them with respect to the recovered model parameters to build an annotated swings database. Then, the system should be able to provide a description of the user's faults by matching his swing parameters against the database.

References

[1] Isard, M. and Blake, A., "CONDENSATION – Conditional Density Propagation for Visual Tracking," *IJCV*, 29, 1, 5–28, Aug. 1998.

[2] I.J. Cox, "A Review of Statistical Data Association Techniques for Motion Correspondence," *Int. J. of Computer Vision*, 10, 1, 53–66, 1993.

[3] M.A Fischler and R.C Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications ACM*, 24, 6, 381–395, 1981.

[4] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 6, 679–697, Nov. 1986.

[5] A. Cochran and J. Stobbs, "Search for the Perfect Swing," *Triumph books*, 1996, ISBN 1-572-43109-1.

[6] Peter J. Rousseeuw and Annick M. Leroy, "Robust Regression and Outlier Detection," *Wiley & Sons*, 1987, ISBN 0-471-85233-3.

[7] Jerry M. Mendel, "Lessons in Digital Estimation Theory," *Prentice-Hall*, 1987, ISBN 0-13-530809-7.

[8] H. Sidenbladh and M. J. Black and L. Sigal, "Implicit Probabilistic Models of Human Motion for Synthesis and Tracking," *ECCV*, 784–800, May 2002.