# A low-cost performance analysis and coaching system for tennis

Philip Kelly, Ciarán Ó Conaire, David Monaghan, Jogile Kuklyte, Damien Connaghan and Noel E O'Connor
CLARITY: Centre for Sensor Web Technologies,
Dublin City University, Ireland

Juan Diego Pérez-Moneo Agapito and Petros Daras
Centre for Research and Technology,
Informatics and Telematics Institute, Greece

## ABSTRACT

In this paper we present an innovative and novel system for tennis performance analysis that allows coaches to review a player's match performance and provide detailed audio-visual feedback to the athlete. The system utilises a simple network of low-cost IP cameras that encompass the tennis court. A graphical user interface provides coaches with video playback feeds from multiple viewpoints, a range of intuitive tools for 2D and 3D annotation, real-time game statistics and the facility for a coach to record audio commentary. This system is specifically designed with non-professional sports clubs in mind, with an emphasis on low-cost equipment. While we focus on tennis in this work, we believe our system can be generalised to a wide range of other sports.

## Categories and Subject Descriptors

I.4.8 [**IMAGE PROCESSING AND COMPUTER VISION**]: Scene Analysis—*Object recognition, Tracking*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

## General Terms

Design, Experimentation, Human Factors

## Keywords

Sports analysis, coaching interface, ball tracking, player tracking, 3D sports visualisation sports feedback

## 1. INTRODUCTION

In order for a coach to improve a player's performance, both technically and tactically, they must be able to ascertain the deficiencies in the athlete's abilities and effectively communicate to the player how to correct these limitations. Traditionally coaches obtain the information required to such make decisions on a sports-person's abilities via statistics obtained from manual, or semi-automatic, annotation entire tennis matches and technical analysis of pre-recorded video clips. For these purposes there are several commercially available solutions, most notably Protracker Tennis [3], ProZone [4] and Dartfish [2]. The primary focus of these systems is to collate data on player performance and to provide statistical feedback to coaches. However, [3] requires significant manual annotation to gain insight into a player's tactical ability, while providing no video data for technical analysis. ProZone [4] performs similar functionality, but in a semi-automatic manner, however it still requires a high level of manual input to correct errors from the automatic processing. The Dartfish system [2] focuses more on analysis of a player's technical ability, providing semi-automated tracking and measurement of a player's bio-mechanical movements, but provides little insight into a player's tactical performance over multiple games.

In this paper, we describe a coaching system that has a number of main objectives; (1) to automate, as much as possible, the manual annotation requirements needed by the coaches – eliminating the laborious overheads of traditional sports coaching systems; (2) to provide the tools that allow a coach to quickly perform a rigorous analysis on the recorded data; (3) to maximise the impact of coaching feedback by providing tools to emphasis and highlight the feedback that they wish to convey to their players; and (4) to provide this coaching functionality by means of a network of low-cost cameras, making these tools affordable to non-professionals.

In Section 2 an overview of the low-cost camera infrastructure is given. Section 3 describes the software components of the system. The performance analysis and visualisation modules are explained in Section 4 and the conclusions are presented in Section 5.

## 2. CAMERA SET-UP AND DATA SET

The dataset used is this paper is from the *3DLife: Sports Activity Analysis in Camera Networks* ACM MultiMedia 2010 Grand Challenge. This dataset includes video streams of competitive singles tennis captured from 9 low-cost cameras, which could feasibly be installed within any local sports club (see Figure 1 for camera network layout). Details of the dataset and camera network set–up can be obtained from the Grand Challenge web site.

## 3. SYSTEM OVERVIEW

Figure 2 shows a flowchart that outlines the proposed coaching system. The video feeds from the dataset are in-

*ACM MultiMedia Grand Challenge* 2010 Florence, Italy

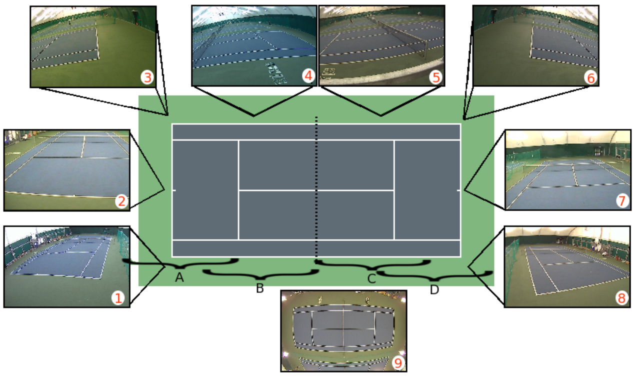


**Figure 1: A schematic layout of a tennis court showing sample images from the nine cameras surrounding the court. Camera number 9 is positioned on the ceiling above the court. The labels A–D will be explained later in Section 3.4.**

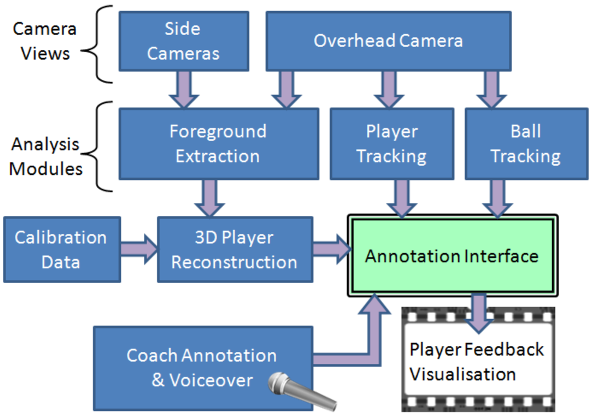
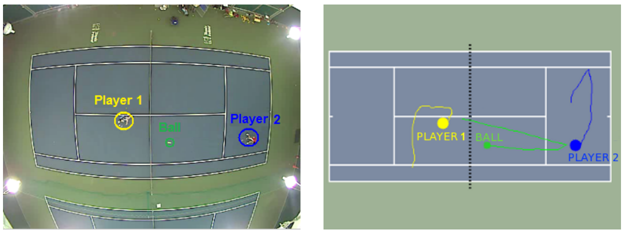


**Figure 2: Overview of our coaching system.**

put to the system software, which consists of three analysis modules, foreground extraction, player tracking and ball tracking. Camera calibration is also carried out using calibration data included in the dataset. The foreground data and camera calibration information are then utilised to perform 3D player reconstruction. These 3D player avatars are generated from all 9 cameras streams and rendered on a 3D virtual tennis court for visualisation, analysis and coach feedback purposes. An annotation interface is also used in the coaching tool and incorporates a built-in record facility, which enables the coach to record an informative voice-over on player performance issues. This can be a valuable tool when the coach does not have direct access to the player.

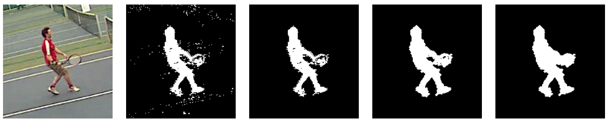
## 3.1 Player and Ball Tracking

In this study, we follow the approach of [12] to provide temporal localisation data on the players and tennis ball. This technique utilises a single overhead camera, which is assumed to be in a fixed position, and tracks the objects in the image plane. These image pixel coordinate locations are subsequently converted into real-world Euclidean coordinates – see Figure 3. In Figure 3 (b) the real world Euclidean positions of the objects are displayed on a 2D virtual tennis court. The median error of these results for the two tracked players was calculated to be 1.10$m$ and 0.99$m$. It should, however, be noted that the ball tracking presented in this paper only determines the 2D position of the ball – the height of the ball above the court is not calculated. In future work we intend to extend the tracking of the ball to 3D by including views from multiple calibrated cameras.



(a) Image Coordinates (b) Euclidean Coordinates

**Figure 3: Example player and ball tracking results.**



(a) (b) (c) (d) (e)

**Figure 4: Foreground extraction post-processing steps: (a) Image; (b) Extracted foreground; (c) Small region removal; (d) Dilation; (e) Holes filled.**

## 3.2 Foreground Extraction

The accuracy of the 3D player reconstruction algorithm, in part, relies on generating accurate athlete silhouettes. To achieve this aim, a layered background model for each camera is generated. Each pixel's model can be represented by up to 5 colours, which are determined by processing an entire video sequence and updating the model with non-motion pixels (determined using 3-frame motion differencing).

At run-time, a pixel is considered to be part of the background if it is within a distance $T$ of any of its background colours within RGB space. Based on our experimental observations we have chosen $T = 20$. As lighting changes and shadows are frequent in the videos, the extracted foreground pixels are then subjected to a further test. A colour/brightness difference value, $D$, is defined as:

$$D = 18 \times D_{gb} + |log(V_{BG}/V_{curr})| \qquad (1)$$

where $D_{gb}$ is the Euclidean distance between the current pixel and the most common background pixel in normalised-$gb$ space; $V_{BG}$ is the background brightness; and $V_{curr}$ is the current pixel brightness. The constant in equation (1) was determined experimentally based on unrelated video data. For a particular foreground pixel, if $D < 0.5$ then that pixel is marked as a lighting change pixel (shadow/highlight) and discarded. The resultant foreground is then subsequently post-processed to remove noise and fill holes in the silhouette as can be seen in Figure 4.

The method was compared with manually annotated ground truth images for 65 results and used the Fuzzy Jaccard index as a comparison metric (the uncertainty tolerance was set to $\sigma = 4$ as proposed by [11]). Using this metric, a value of 1.0

indicates perfect segmentation. Table 1 shows the results from this comparison and shows that the technique used in this paper out-performed the standard mixture-of-Gaussians (MoG) model and frame differencing.

| Method | Results A | Results B |
|---|---|---|
| Layered | 0.10 | 0.74 |
| MoG | 0.60 | 0.58 |
| Frame differencing | 0.34 | 0.42 |

**Table 1: Foreground segmentation accuracy results (Fuzzy Jaccard index). Results are shown (A) before post-processing and (B) after post-processing.**

## 3.3 Camera Calibration and Synchronisation

The Grand Challenge dataset video sequences were accompanied by calibration data, however they were not synchronised. The Matlab Camera Calibration Tool [1] was used to calculate the intrinsic parameters and the OpenCV camera calibration toolbox was used to calculate the extrinsic parameters of the cameras. To account for synchronisation a camera-sync algorithm [10] was adapted to align the input foreground silhouettes. The mean back projection error of the 3D points, provided in the challenge dataset, used in order to calibrate the extrinsic parameters was 3.33777 pixels. This is a relatively high error and is due, mainly, to the high distortion and low resolution of the cameras.

## 3.4 3D Player Reconstruction

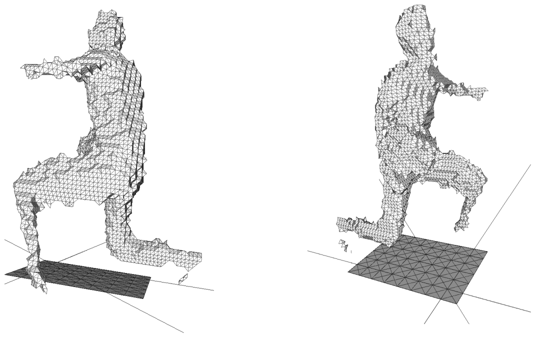
There are a variety of 3D reconstruction methods presented in the literature [6, 13, 7]. In this paper, a volumetric intersection technique is employed to perform 3D reconstruction. Using this technique the visual hull of the 3D object is computed, which is based on the shape-from-silhouette method presented in [5]. The visual hull is created from intersection cone projections obtained from each camera view. Each cone projection takes into account both the silhouette and the rays between the center of the camera and the contour of the silhouette.

One of the main advantages of this method is that no texture correspondences are needed and therefore it can produce high quality 3D reconstructions from poor quality video inputs. In the Grand Challenge dataset the tennis players are not captured by all the cameras simultaneously and thus cannot be used for calculating the visual hulls of the players. To account for this we have divided the tennis court into 4 sections (A, B, C, D – see Figure 1) and the required visual hulls are computed from the corresponding section.

Following from [10] let $\mathbf{I}^c$ be the image projection captured by a camera $c$, where $c = 1, ..., N$ and $N$ is the total number of cameras. $\mathbf{I}^c$ is projected onto a base plane, $z_0 = 0$, forming $\mathbf{I}^c_z$. All other planes, $z_s$, are parallel to $z_0$ and occur at distances $d(z_{s+1} - z_s) = \rho =$ const, where $s = 0, ..., K$ and K is the total number of planes per camera. Let $W$ be the width and $H$ be the height of the tennis court then the dimension of $\mathbf{I}^c_z$ is defined to be $(\frac{W}{\rho} \times \frac{H}{\rho})$. Let $\mathbf{S}^c_z$ be the silhouette projection in a plane $z$ captured by a camera $c$ and $\mathbf{S}^c_z \subset \mathbf{I}^c_z$. Therefore,

$$\mathbf{I}^c_z(i,j) \in \mathbf{S}^c_z \; if \; and \; only \; if \; \mathbf{I}^c_z(i,j) = 1$$

where $i = \dfrac{x}{\rho}$ and $j = \dfrac{y}{\rho}$ and $x, y$ are the coordinates of a point that lies on a plane $z_s$.



(a) 3D Model View 1

(b) 3D Model View 2

**Figure 5: Example 3D player reconstruction.**

The Visual Hull is obtained by the intersection of planes from all cameras. If a 3D point is simultaneously captured by all cameras of a section for a specific plane then that point belongs to the Visual Hull. If $n \in N$ is the total number of cameras in a section then the intersection matrix $\mathbf{M}_z$ can be defined as follows:

$$\mathbf{M}_z = \sum_{c=1}^{n} \mathbf{I}^c_z$$

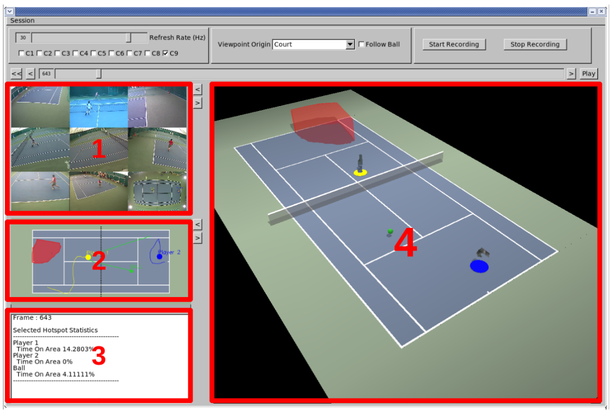
The 3D mesh is subsequently calculated using the marching cubes algorithm [9] where $N = 9$, $\rho = 0.15$ meters and $K = 12$ for each camera. An example of 3D player reconstruction can be seen in Figure 5.

# 4. VISUALISATION AND ANNOTATION

The user interface of the proposed coaching system displays the input video, annotation tools, output statistics and a 3D view (labeled 1-4 respectively in Figure 6), with an intuitive browser on the top. The user can quickly surf an entire game via the browser, pause, step forward/backward, or play back at a chosen frame rate. In the input video pane the user can designate what cameras to view and the tool will automatically scale and position the camera images to make efficient use of screen real-estate. The user can also annotate the images in this pane in order to highlight points of interest to the athlete – see Figure 7.

The annotation pane, labeled (2) in Figure 6, depicts the tracked ball and player positions throughout the match (as seen in Figure 3). There are two types of annotations are available in this interface. The first are hotspots [8] (see Figure 8 (a)) and the second are player annotations (see Figure 8 (b)). Hotspot annotations are selected 3D regions of interest that have been created from the 2D plan-view of the court. Statistics (e.g. time spent by a player or ball in that area of the court, average velocity through the area, etc.) are calculated for each hotspot. Player annotations allow the coach to highlight suggested adjustments of a player's movement about the court. In Figure 8 (b) the real movements of player 1 are highlighted in red and the coach's annotation are depicted in green and depict how the athlete should have moved during that particular sequence of play.

All the annotations and tracking data are rendered in real-time in the 3D pane, which allows the game to be viewed from any angle, enabling the coach to further clarify an issue



30 Refresh Rate (Hz)
□ C1 □ C2 □ C3 □ C4 □ C5 □ C6 □ C7 □ C8 ☑ C9
Viewpoint Origin Court
□ Follow Ball
643

```
Frame : 643

Selected Hotspot Statistics
--------------------------------
Player 1
  Time On Area 14.2803%
Player 2
  Time On Area 0%
Ball
  Time On Area 4.11111%
--------------------------------
```

**Figure 6: Annotation coaching tool.**

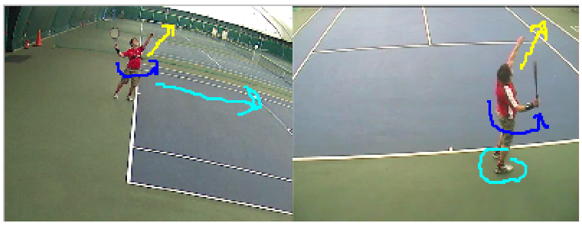


**Figure 7: Input Image Annotation.**

to a player. The coach can also replay the match from the point of view of either player or from the sideline – see Figure 9. This feature allows the player to relive the match from their opponent's point of view own viewpoint, or a spectator's viewpoint, to aid in the post-match analysis of tactical performance. In this mode, the 3D animation follows the course of the ball when it is in play, similar to how a player will closely watch the ball in a match, adding a further level of realism to the system.

## 5. CONCLUSIONS

In this paper we presented a novel tennis analysis system to review a player's match performance and provide detailed feedback to the athlete via 3D reconstruction and audio annotation. The system provides a range of intuitive tools including 2D and 3D annotation, video playback from multiple viewpoints, real-time game statistics and the facility to record audio commentary. This framework is specifically designed for non-professional sports clubs and uses low-cost equipment. The application presented in this work is targeted at tennis but the framework can be utilised by a wide range sports. In future work, we wish to perform usability studies with a large number of coaches and to improve its performance based on the feedback from these studies.

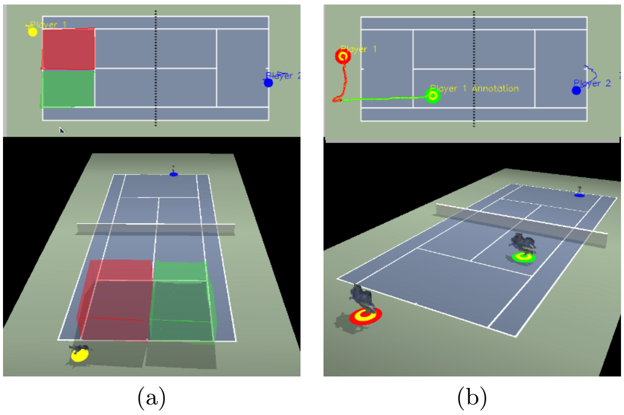


**Figure 8: (a) Hotspot Annotations; (b) Player Annotations.**

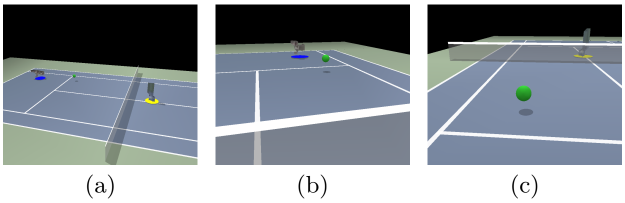


**Figure 9: (a) View from the crowd; (b) Player 1 viewpoint; (c) Player 2 viewpoint.**

## Acknowledgments

**This work is supported by Science Foundation Ireland under grant 07/CE/I1147. This research was partially supported by the European Commission under contract FP7-247688 3DLife.**

## 6. REFERENCES

[1] Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/.
[2] Dartfish. http://www.dartfish.com/.
[3] Protracker tennis. http://www.fieldtown.co.uk/.
[4] Prozone. http://www.prozonesports.com/.
[5] S. Chang and Y. Wang. Three-dimensional object reconstruction from orthogonal projections. 7(4):167–176, December 1975.
[6] R. Hartley. In defence of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6), June 1997.
[7] R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision*. Ed. Cambridge, 2000.
[8] P. Kelly and N. O'Connor. Vision-based analysis of pedestrian traffic data. In *CBMI*, pages 133–140, 2008.
[9] W. Lorensen and H. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.
[10] T. Matsuyama and [et al.]. Real-time dynamic 3-d object shape reconstruction, and high-fidelity texture mapping for 3-d video. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3), 3 2004.
[11] K. McGuinness and N. E. O'Connor. A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition*, 43(2):434–444, February 2010.
[12] C. Ó Conaire, P. Kelly, D. Connaghan, and N. E. O'Connor. Tennissense: A platform for extracting semantic information from multi-camera tennis data. In *International Conference on Digital Signal Processing (DSP)*, 2009.
[13] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans PAMI.*, 15(4):353–363, Apr 1993.