Human Tracking in Sports Video

By: Terence Bloom
Supervisor: Dr. Andrew Bradley

Context

It is the problem of getting a computer to be able to analyse a digital video stream and interpret it visually as a scene, such that it is able to detect when a person is present in the scene, and subsequently track the person's movement through the scene. A computer being able to track humans would be useful for a wide range of applications from automated tracking security or TV cameras, to a vision-based human-computer interface. A different type of application for this technology is to automate the task of annotating digital video with metadata that denotes the presence and describes the actions of people in the video. Annotating digital tennis footage with metadata such as the time codes of strokes played would provide access and search facilities on a tennis footage video archive. A tennis coach, for example, would benefit from this by being able to easily retrieve, and view straight from a computer, footage of certain strokes of his protégé to track their improvement over time, or perhaps of an up-coming opponent for analysis to identify their weaknesses for exploitation.

AIM

The aim of this thesis was to develop a system that is able to track at least the forecourt player as they move around the tennis court, by knowing at all times whether the player is present, and drawing a bounding box around them while they are present. Furthermore, for instances in time which are known to correspond to the play of a stroke, the system should perform further analysis to make an elementary attempt at recognising the stroke. This aim is the tactical precursor to the automatic annotation of digital tennis footage.

Method

The algorithm developed for achieving this aim is divided into three units:

1. Finding, through image processing techniques, the connected component in a given frame that is probably the forecourt player.
2. Tracking the player from one frame to the next by coordinating, and utilising the results of the visual processing for each frame.
3. Recognising the player’s strokes in key frames that correspond to the play of a stroke, using derived information about the different tennis strokes.

i. Start with a reference image of the court without any players present, and a new frame from the footage.
   ii. Convert both images to greyscale, then low-pass filter them both to remove random noise and detail.
   iii. Take the pixel-by-pixel absolute value difference.
   iv. The difference image's histogram exhibits a large dominant mode near zero (representing the static background pixels), and a smaller mode at a higher level (mostly representing the variant player pixels).
   v. Threshold the difference image to separate the two modes, using a global threshold value found from the following equation:
      \[ \text{threshold} = \text{mean} + \frac{\text{standard deviation}}{\text{PSI}} \]
      where 'PSI' (Player Size Index) constant is assigned appropriately, based on the player's size in the previous frame (as their size should be very similar in this frame), such that the required percentage of pixels will be designated white after thresholding the current frame.
   vi. Use standard morphological filtering methods to reduce noise in the binary difference image.
   vii. Iterate through the remaining connected components (blobs) to find one that is large enough, and the correct dimensions to be human.
   Use colour matching if there is more than one possible blob.
   viii. The midpoint of the bounding box of the player found is considered a good representation of their position.
   iii. This has to be confirmed as the forecourt player by checking their midpoint hasn’t moved unrealistically between consecutive frames.
   iv. A flag is maintained indicating whether the player is expected to be present. It is only changed if the player ‘disappears from, or ‘appears’ into view.
   v. As a first attempt, look for a blob that could represent the racquet and ball's position. It is required to be big enough, the correct dimensions, a realistic distance from the player, and each colour channel within an allowable range.
   vi. If one is found, its relative position to the player is analysed. Simple heuristics are utilised to recognise the stroke: looking at a right-handed player from behind, if the racquet is on the right its is a forehand, the left is a backhand, and above is a serve or smash.
   vii. If one wasn’t found, proceed to skeletonise the player’s blob and analyse it, using similar heuristics to decide on the stroke.