Real Time Tracking for Enhanced Tennis Broadcasts

Gopal Sarma Pingali, Yves Jean and Ingrid Carlbom
Visual Communications Research Department
Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974
{gsp,yvesjean,carlbom}@research.bell-labs.com

Abstract

This paper develops real time tracking technology for sports broadcast applications. The specific sport chosen here is the game of tennis. The outputs of the tennis tracking system are spatio-temporal trajectories of motion of the players and the ball which can in turn provide a number of statistics about the game. For instance, the distance travelled by a player, the speed and the acceleration at any instant, as well as court coverage patterns can be obtained from the trajectories. The statistics so obtained can be visualized in compelling ways to enhance the appreciation of the athleticism and strategy involved in the sport. We present techniques for tracking the players and the ball in video obtained from stationary cameras. The problem is challenging as the tracking needs to be performed outdoors, players are fast-moving non-rigid objects, and the ball is a small object that can move at speeds in the range of 150 miles an hour. Player trajectories are obtained by dynamically clustering tracks of local features. Ball segmentation and tracking is based on shape and color features of the ball. Real time tracking results are presented on video recorded live by the authors in an international tennis tournament.

1 Introduction

Computer vision is increasingly being used as a tool for capturing and modeling live action in a variety of applications such as visual surveillance [3, 16, 4, 17, 14, 21, 7, 13, 20]. Sports present a useful and challenging area for the application of computer vision technology for several reasons. Most sports involve complex human motion and, therefore, capturing, analyzing, quantifying and highlighting the ability of the athlete can help audiences better appreciate the sport. While capturing general human motion is hard, sports can be favorable to the application of computer vision due to the rigid rules defining a sport and availability of well-defined landmarks. Sports can still be a very challenging computer vision problem as they typically involve fast and complicated actions. Sports are a very popular form of entertainment and routinely involve a tremendous amount of technology, particularly video technology. A successful application of computer vision to sports could reach millions of television viewers and web surfers. Application of computer vision to sports can range from real-time tracking to the capture of complete three dimensional models. A computer vision system can be a critical component of a larger system for presenting a sport.

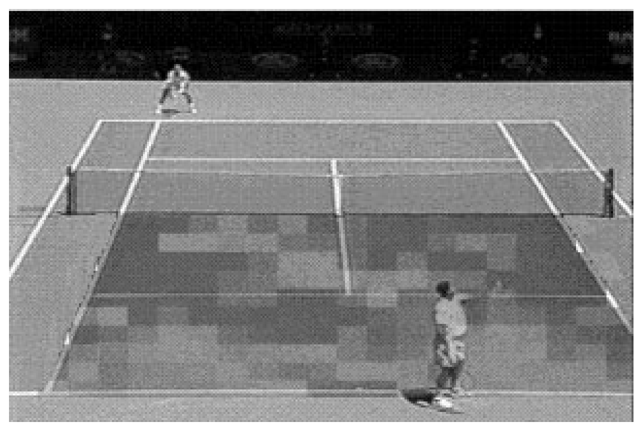

Figure 1. Color-coded court occupancy map showing time spent by the player in different areas of the court; light areas (red, light green) indicate high occupancy while dark areas(dark green, blue) indicate low occupancy

In this paper, we develop real-time tracking techniques for enhancing broadcasts of tennis matches. A tennis tracking system should provide spatio-temporal trajectories of the players and the ball in real time. Such spatio-temporal trajectories can provide a wealth of information about the game. The trajectories can be the basis for obtaining information hitherto unavailable such as the distance traveled by the player over the course of a game, a set or a match, the instantaneous, average, and peak speed and acceleration of a player, the areas of the court covered by a player, and the ball placement strategy of the player. Such information can be presented instantaneously or archived for reference in future games. Figure 1 shows a color coded court occupancy map which indicates the amount of time spent by the player in different parts of the court, as obtained from the motion trajectories of the player. This is an example of how the trajectories can be used to visualize the game strategy and player strengths and weaknesses.

2 Related Research

There has been considerable work on tracking people over the years including early work on moving light displays [15] and several camera based systems for tracking moving people (e.g., [11, 2, 12, 6, 3, 16, 19, 17, 14, 21, 7, 13, 20]). A recent review of work in this area is provided in [1]. In spite of the increased interest in tracking people in recent years, very few systems have been reported that can track people in real time. The authors are not aware of any reported work on tracking people in real time in outdoor situations such as the problem addressed in this paper. Color based segmentation techniques such as those used for tennis ball segmentation in this paper have been addressed by a number of authors in the past (e.g.,[9, 10]). In this work we study the color characteristics of a tennis ball and develop a fast algorithm for ball segmentation and tracking.

3 Tracking Player Motion

The outputs we desire of a player tracking system are trajectories, one per player, showing in real time the movement of player during the game. This is challenging as it is hard to obtain a clean segmentation of the player at all times. Differentiating the player from the background, especially in real time, is complicated by changing lighting conditions, wide variations in clothing worn by players, differences in visual characteristics of different courts, and the fast and non-rigid motion of the player. The central problem is that real time segmentation techniques do not yield a single region or a consistent set of regions as the player moves across the court. In order to robustly obtain the player trajectories, we track local features and derive the player trajectory by dynamically clustering the paths of local features over a large number of frames based on consistency of velocity and bounds on player dimensions.

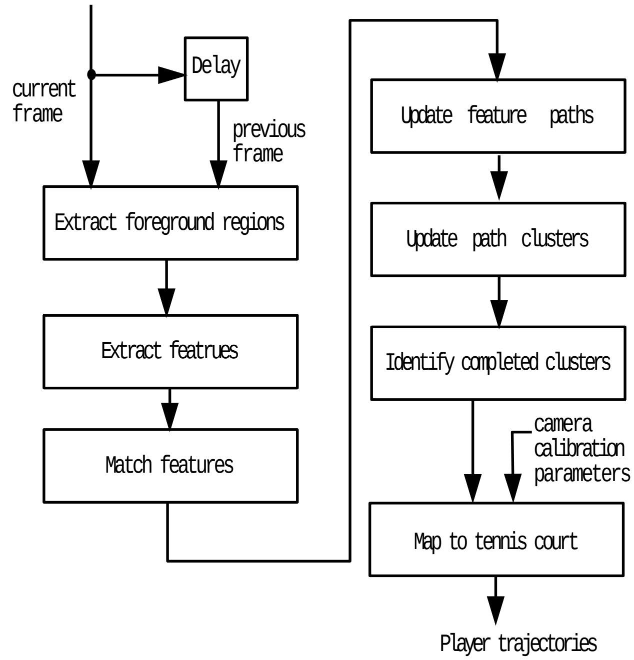

Figure 2. Steps in tracking player motion from video

Figure 2 summarizes the steps involved in the player tracking system. Differencing of consecutive frames followed by thresholding is used to extract the regions of motion as this is a fast operation and works across varying lighting conditions. Small gaps in the extracted motion regions are filled by performing a morphological closing operation [8]. Thus,

$$B_t = (H_T(I_t - I_{t-1}) \oplus g) \ominus g \qquad (1)$$

where B_t is a binary image consisting of regions of interest at time t, I_t is the input image at time t, H_T is a thresholding operation with threshold T, g is a small circular structuring element, and $\oplus$, $\ominus$ indicate morphological dilation and erosion operations respectively. Consistent segmentation of a moving player is not obtained even after this operation. The number of regions per player change in shape, size and number across frames.

In order to obtain player trajectories, local features on the extracted regions are identified in each frame. Extrema of curvature on the bounding contours of the regions were chosen as the local features. Features are matched across frames by minimizing a distance given by $k_r \delta r^2 + k_\theta \delta\theta^2 + k_\kappa \delta\kappa^2$ where δr is the euclidean distance between feature positions, $\delta\theta$ is the difference in orientation of the contours at the feature locations, $\delta\kappa$ is the difference in curvature of the contours at the feature locations and k_r, k_θ, k_κ are weighting factors. A sequence of feature matches is considered a feature path that indicates the motion of a feature over time and is given by $\Phi(\mathbf{x}, \mathbf{y}, \mathbf{t}, l, \mu_x, \mu_y, \sigma_x, \sigma_y)$ where $\mathbf{x}, \mathbf{y}, \mathbf{t}$ are vectors giving the spatio temporal coordinates at each instant, l is the temporal length of the path, and μ_x, μ_y are the mean x and y values over the path and σ_x, σ_y are the variances in x and y values over the path.

Feature paths having sufficient temporal overlap are dynamically merged in each frame to form clusters. Cluster merging is also performed in a similar fashion. A cluster is given by , $(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{f}, l, p, \mu_x, \mu_y, \sigma_x, \sigma_y)$ where $\mathbf{f}$ gives the number of features contributing to a cluster at any instant, p is the total number of paths contributing to the cluster, (μ_x, μ_y) indicate the mean displacement of contributing features from the cluster coordinates and (σ_x, σ_y) indicate the variance in displacements. Two clusters or a path and a cluster that are close are merged based on a distance measure given by $\lambda_x \Delta\sigma_x + \lambda_y \Delta\sigma_y + \lambda_\tau \Delta\tau$ where $\Delta\sigma_x, \Delta\sigma_y$ are the maximum change in variances of x and y displacements of features resulting from merging the clusters, $\Delta\tau$ is the normalized squared sum of the difference in orientations of the velocity vectors along the trajectories corresponding to the two clusters, and $\lambda_x, \lambda_y, \lambda_\tau$ are weighting factors based on bounds on the size of a player.

The clustering algorithm is capable of tracking several people in real time. Clustering can also be optimized given the expected number of objects in the field of view of a camera for the tennis application. An existing cluster of sufficient temporal length is extended based on its last position when there is negligible motion in the scene in order to continue tracking a still player. Trajectories are mapped on to the court ground plane using camera calibration parameters [18].

Figure 3 shows an example of the break-up of tracking steps for a single frame in a tennis video. The image on the left shows the original frame while the frame on the right shows the detected contours after morphological closing, positions of features on the contours, the feature paths in the current frame and the updated object trajectory at that instant.

4 Tracking Ball Motion

To follow the action in the game of tennis, the motion of the ball also needs to be tracked. This can be challenging as the the ball is a small object (diameter approximately 6.5 cm) (see for e.g., [5]) travelling at speeds up to 150 miles per hour (approximately 67 ms^{-1}) over a court which is approximately 24m long and 11m wide in which moving players are also present. Hence, tracking the motion of the ball across the court demands processing of video at high frame rates and coverage of the court with multiple cameras. We first focused on a high-speed technique for ball segmentation and tracking.

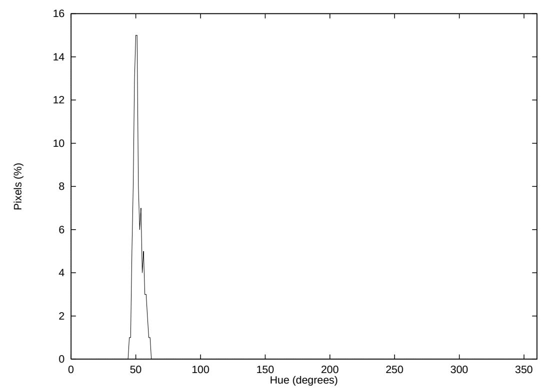

Figure 4. Distribution of tennis ball hue in images taken under different lighting conditions

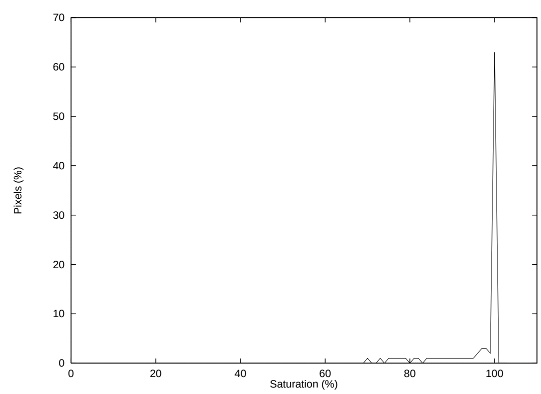

Figure 5. Distribution of tennis ball saturation in images taken under different lighting conditions

A significant cue for tennis ball segmentation is the standardized yellow color of balls used in most tennis tournaments. Figure 4 shows the distribution of hue (in HSV space) of pixels corresponding to a tennis ball obtained from several images of the ball taken under varying outdoor lighting conditions with a color camera positioned at different distances from the ball. It is seen that the hue distribution has a sharp peak at about 50 degrees. The segmentation can be made more robust

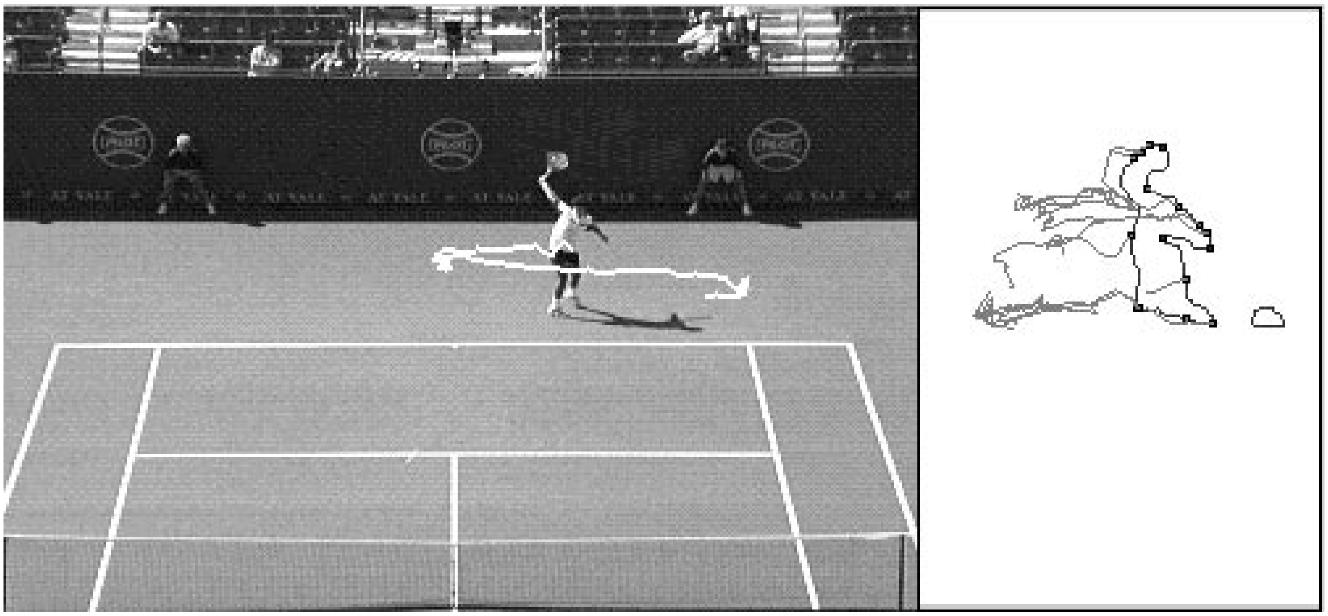

Figure 3. Contours and feature paths obtained in player tracking with stationary camera

by taking advantage of the distribution (figure 5) of saturation values of tennis balls as well.

Figure 6 shows the steps in the algorithm we have developed for tracking the motion of a tennis ball. Motion regions are first identified by subtracting the previous frame from the current frame and thresholding the result. A morphological closing operation is then performed to close gaps in the binary image as in the case of player tracking. At this stage, the ball region is separated from other moving objects such as players by using the color based segmentation. Regions of the current frame corresponding to the filtered motion regions are converted to HSV space and further filtered based on the hue-saturation model of the tennis ball. The region with highest confidence based on color and size is chosen as the ball region. The center of the ball region is then matched with the location in the previous frame to update the ball trajectory. The algorithm is very efficient as the conversion to HSV space and color based segmentation are applied only to motion regions in the image. The implemented algorithm runs at video rate.

5 Experimental Results

The player tracking system has been tested on over five hours of video recorded by the authors by means of color cameras placed in a stadium during an international tennis tournament. The recording was done over different times of day as well as night (under floodlight illumination). To achieve higher tracking accuracy, four camera positions were used – each covering one half of the court. Video was obtained with two cameras at the ends of the court (facing the baseline) and with two cameras on one side (facing the sidelines). The player tracking algorithm runs at approximately 20 frames a second on an SGI Indigo workstation. The algorithm was tested both by running it for several hours on live input coming directly from video tapes and observing the output trajectories and by running it on several 300 second clips digitized from different portions of the video tapes and carefully analyzing the performance of the algorithm. The algorithm was found to be reliably tracking player motion on video recorded in different lighting conditions. Figure 7 shows a real-time tracking sequence for one of the camera views with the camera facing the baseline.

The ball tracking algorithm could not be tested on the video recorded in the tournament as the shutter speed used for the recording was too low. The implemented ball tracking algorithm has been tested on test sequences in which the authors hit tennis balls with tennis racquets. Ball segmentation and tracking results on these initial sequences are very encouraging. The ball speeds for these sequences were in the range of 60 miles an hour. Figure 8 shows an example of ball segmentation while figure 9 shows the ball trajectory obtained on one sequence with the algorithm running at 30 frames a second.

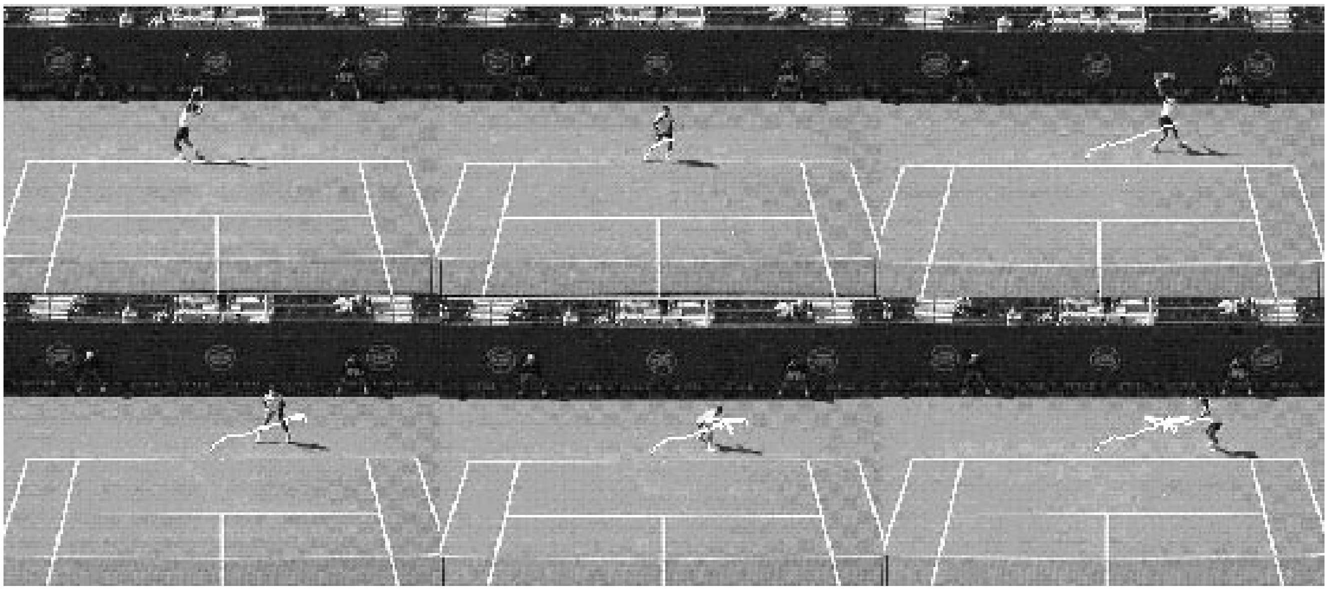

Figure 7. Sequence of tracking player over 12 seconds during a point

6 Conclusions and Further Research

This paper presented real-time algorithms for tracking player and ball motion in tennis games. This is an example of computer vision technology for analyzing sports action and obtaining a rich set of game statistics. The real-time results obtained on several hours of tennis video are promising. Future research should address the issues of tracking ball motion across multiple cameras, optimal camera placement and stereo for 3D ball localization. Future challenges also include real time techniques for tracking the motion of a racquet, the joint angles of a player, as well as extension to other sports.

References

[1] J.K. Aggarwal and Q. Cai. Human motion analysis: A review. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 90–102, 1997.

[2] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, 17:73–83, 1984.

[3] Q. Cai and J.K. Aggarwal. Tracking human motion using multiple cameras. In *International Conference on Pattern Recognition*, pages 68–72, 1996.

[4] Shankar Chatterjee, Ramesh Jain, Arun Katkere, Patrick Kelly, Don Y. Kuramura, and Saied Moezzi. Modeling and interactivity in *mpi-video*. Technical Report VCL-94-103, Visual Computing Lab, Univ. of California, San Diego, 1994.

[5] Paul Douglas. *The Handbook of Tennis*. Alfred A. Knopf, New York, 1996.

[6] Kazuhiro Fukui, Hiroaki Nakai, and Yoshinori Kuno. Multiple object tracking system with three level continuous processes. In *IEEE Workshop on Applications of Computer Vision*, pages 19–27, 1992.

[7] D.M. Gavrila and L.S. Davis. Towards 3-d model-based tracking of humans in action. In K. Bowyer and N. Ahuja, editors, *Advances in Image Understanding, A Festschrift for Azriel Rosenfeld*, pages 264–279. IEEE Computer Society Press, 1996.

[8] C. R. Giardina and E. R. Dougherty. *Morphological Methods in Image and Signal Processing*. Prentice Hall, 1988.

[9] G. Healey, S. Shafer, and L. Wolff, editors. *Physics-Based Vision: Principles and Practice, COLOR*. JOnes and Bartlett, Boston, 1992.

[10] G. Healey and D. Slater. Global color constancy: recognition of objects by use of illumination invariant properties of color distributions. *Journal of Optical Society of America A*, 11(11):3003–3010, November 1994.

[11] D. Hogg. Model based vision: A program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

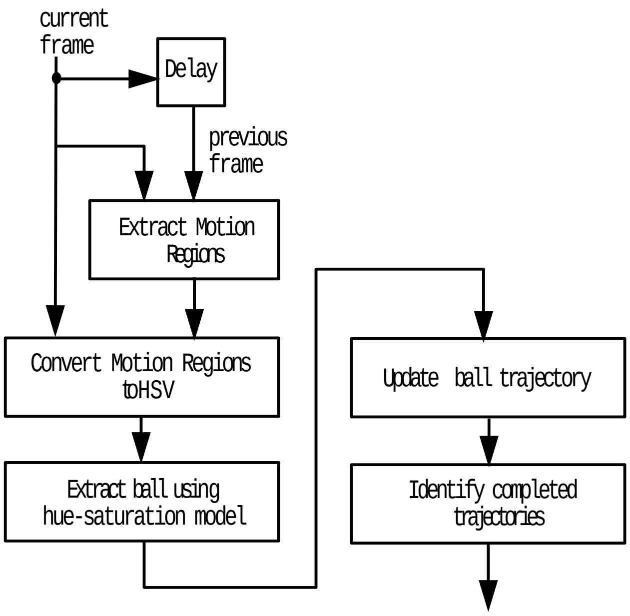

Figure 6. Steps in tracking ball motion from video

[12] S. Intille and A. Bobick. Visual tracking using closed worlds. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 672–678, 1995.

[13] I.A. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 81–87, 1996.

[14] Sarma Pingali and Jakub Segen. Performance evaluation of people tracking systems. In *Third IEEE Workshop on Applications of Computer Vision*, 1996.

[15] R.F. Rashid. Towards a system for the interpretation of moving light displays. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:574–581, February 1980.

[16] K. Rohr. Towards model based recognition of human movements in image sequences. *Computer Vision Graphics and Image Processing:Image Understanding*, 59(1):94–115, January 1994.

[17] Jakub Segen and Sarma Pingali. A camera based system for tracking people in real time. In *IAPR International Conference on Pattern Recognition*, pages 63–67, 1996.

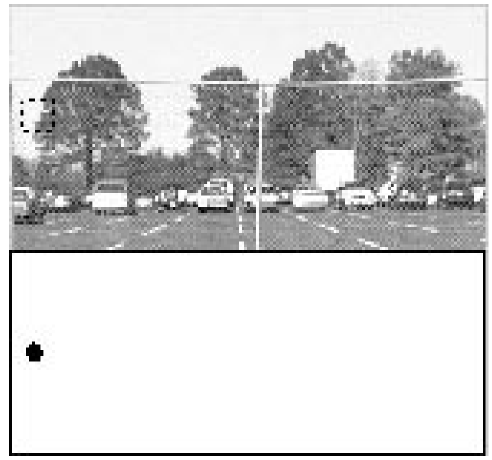

Figure 8. Example ball segmentation

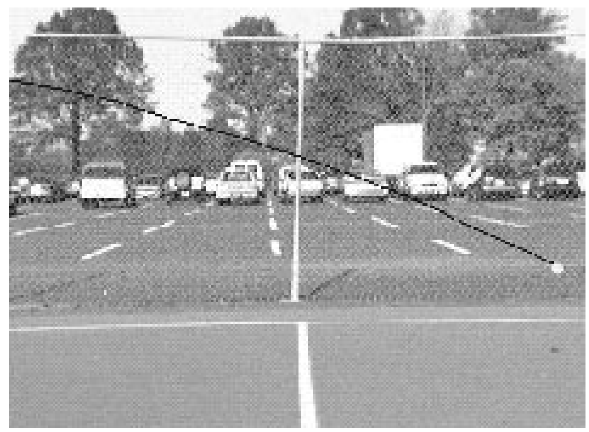

Figure 9. Example ball tracking over 500ms

[18] R.Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[19] V.T. Tsukuyama and Y. Shirai. Detection of the movements of persons from a sparse sequence of tv images. *Pattern Recognition*, 18:207–213, 1985.

[20] S. Wachter and H.-H. Nagel. Tracking of persons in monocular image sequences. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 2–9, 1997.

[21] C. Wren, A. Azerbayejani, T. Darrell, and A. Pentland. Pfinder: Real time tracking of the human body. In *Proceedings of SPIE*, volume 2615, pages 89–98, 1996.