Player Detection and Tracking in Broadcast Tennis Video

Conference Paper · January 2009

DOI: 10.1007/978-3-540-92957-4_66 · Source: DBLP

Kuan-Ting Lai
National Taiwan University

Chaur-Heh Hsieh
Ming Chuan University

Player Detection and Tracking in Broadcast Tennis Video

Yao-Chuan Jiang[1], Kuan-Ting Lai[2], Chaur-Heh Hsieh[3], and Mau-Fu Lai[4]

[1] I-Shou University, Kaohsiung County, Taiwan, R.O.C.
m9503023@stmail.isu.edu.tw
[2] Quanta Research Institute, Taoyuan, Taiwan, R.O.C.
kuantinglai@gmail.com
[3] Ming-Chuan University, Taoyuan, Taiwan, R.O.C.
hsiehch@mail.mcu.edu.tw
[4] Tungnan University, Taipei County, Taiwan, R.O.C.
lai@mail.tnu.edu.tw

Abstract. In this paper, we propose a novel algorithm for player detection and tracking in tennis games. The algorithm utilizes court knowledge as well as player color and edge information to extract deformable player figures. Several new techniques are presented in our algorithm: initially, the court lines are detected and reconstructed. Based on the court model, an adaptive search window is designed for locating the minimum region containing a player figure. After retrieving the region of interest, pixel data are processed by non-dominant color extraction and edge detection filters, respectively. Finally, the non-dominant color map and edge map are refined and combined, and a novel shadow removal method is then applied to isolate the player figure. The algorithm was tested on numerous videos with different courts and light condition. Experiments reveal promising results against various environmental factors.

1 Introduction

In recent years, sport video automatic annotation has attracted many research interests. Among numerous research domains, player detection and tracking is a fundamental but also most challenging area. A robust detection and tracking algorithm is required for many high-level operations such as player action recognition or content classification. Many relevant works of player detection have been published in past years [1-3, 5-11]. Early works explore temporal information of frame difference and then perform morphology operations [3, 10, 11] to extract player figure. Those methods are simple and fast, but easily affected by spectator movement or camera view change. Another approach is background subtraction, which constructs a background model to separate players [1, 2, 5, 6, 7, 9]. Major background models include empty court image, mean of continuous frames, and mean of dominant color. The empty court image is hard to retrieve and thus unrealistic. Using continuous frames to set up the statistical background model

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 759–770, 2009.

shows great performance at fixed camera view, but not suitable for circumstance of frequently changed perspectives. The dominant color method has merits of computation simplicity and robustness under different perspectives. However, dominant color selection and range determination are still open problems requiring more effort. Furthermore, in tracking of players, the existing algorithms often employ a search window or bounding box that do not provide a close fit to a player's body. Although it makes no difference in tracking players, high level operations, like action recognition, still demand a best fit window and a complete body. In this paper, we propose a new player detection method which associates non-dominant color extraction and edge detection to effectively separate players from background. Moreover, an adaptive search window and varying bounding box are designed to make the extraction of player body more complete and the tracking more efficient and reliable.

The essential elements of our algorithm can be summarized as three blocks: adaptive search window, non-dominant color extraction and edge detection filter, and player refinement. The adaptive search window based on court knowledge is applied for locating the minimum region that contains a player figure. The region of interest is taken as input for non-dominant color extraction and edge detection filters. The two output maps are then combined to achieve the final detection result. The paper is organized as follows: Section 2 introduces the fundamental algorithms of the proposed system, including adaptive search window, non-dominant color extraction, edge detection and player refinement flow; Section 3 demonstrates experimental results of different tennis games, and the conclusion is given in Section 4.

2 Player Detection and Tracking System

The proposed player detection and tracking flow is illustrated in Fig. 1. For each input frame, we need to detect court lines and build the court model by using the work presented in [12], which is briefly described as follows. Court lines are white and can be detected by extracting white pixels. Nevertheless, the intensities of white pixels are changed by the weather, camera angle, different courts, etc. An adaptive threshold scheme is presented for adjusting color value of court lines. Moreover, some court lines often disappear during the zooming and panning of video camera. For missed lines, an algorithm is derived to efficiently reconstruct the court. The reconstructed court model is used to determine a search window that contains the player figure. The search window determines the region of interest (ROI) to be processed. The initial search window is fixed, whereas the search windows in the subsequent frames are adaptive (varying). The data in ROI are fed into the player detection unit which combines non-dominant color extraction with edge detection to extract player information. The detected result is further refined to achieve a complete player figure. The details are described in the following subsections.

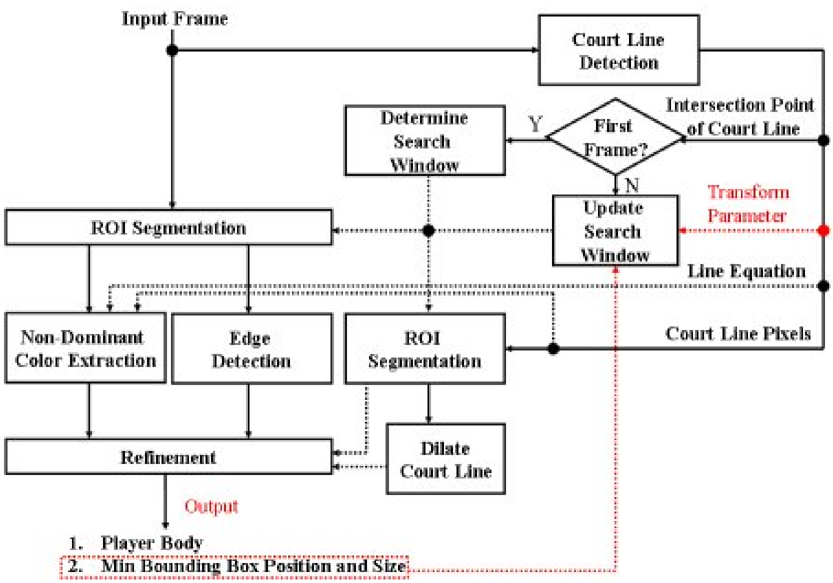


Fig. 1. Flow chart of player detection and tracking

2.1 Adaptive Search Window

At the first frame, the player position is unknown. By referring to the court model shown in Fig. 2, we define initial search areas around the court, where (X_{Pi}, Y_{Pi}) denotes the coordinate of a point Pi. The search areas contain upper court and lower court, as defined by the following equations.

$$UpperCourt$$

$$Left: \left[X_{P4} - \frac{1}{2} * (X_{P4} - X_{P1})\right] or(zero)$$

$$Right: \left[X_{P20} + \frac{1}{2} * (X_{17} - X_{P20})\right] or(image\_width)$$

$$Top: \left[max(Y_{P4}, Y_{P20}) - \frac{2}{3} * max(Y_{P4}, Y_{P20})\right]$$

$$Bottom: [Y_{PC}]$$

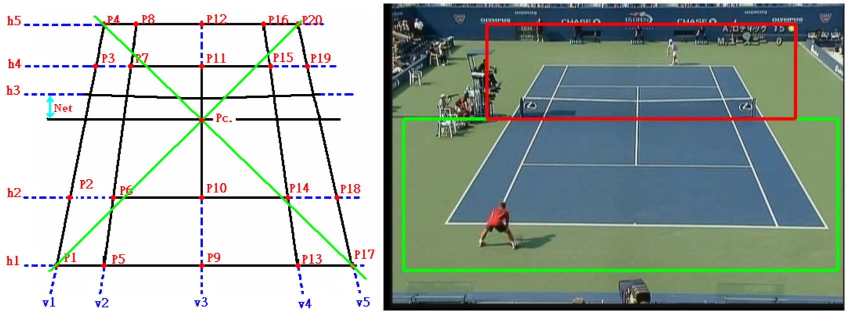
$$LowerCourt$$

$$Left: \left[X_{P1} - \frac{2}{3} * X_{P1}\right] or(zero)$$

$$Right: \left[X_{P17} - \frac{2}{3} * (image\_width - X_{P17})\right] or(image\_width)$$

$$Top: [Y_{PC}]$$

$$Bottom : \left[max(Y_{P1}, Y_{P17}) - \frac{2}{3} * (image\_height - max(Y_{P1}, Y_{P17})) \right]$$



(a) (b)

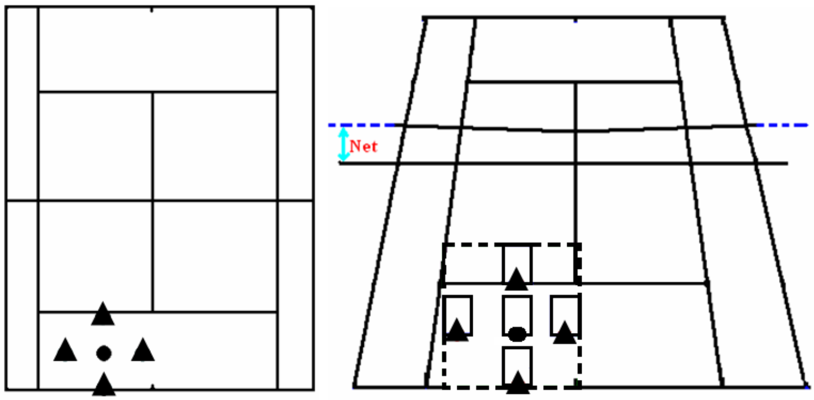
Fig. 2. (a) Court line model and (b) Initial search window

The initial search window can be used to locate the player in a video. Since a player is not a rigid object, we propose an adaptive search window to efficiently track the deformable player figure. According to [5], the speed of a player is around 2 - 7 meters per second. As a result, we use the maximum speed of 7 meters divided by the frame rate as the definition of the search window. Since the speed is only true in the real court, we apply the perspective transform to relate the coordinate in image space to that in real world. The generated adaptive search window is illustrated in Fig. 3 and the procedures are described below:

1. In image space, detect a player and calculate its centroid (c_x, c_y)
2. Map the centroid back into real court model (r_x, r_y) by perspective transform
3. Calculate the maximal possible displaced locations in 4 directions (left, right, up, down), as the arrows shown in Fig. 3 (a)
4. Map the four locations in real court model into image space using perspective transform, as shown the arrows in Fig. 3 (b). The resulting locations indicate the possible centroids of a player in image space
5. Each possible centroid corresponds to a minimum bounding box. Using the minimum bounding boxes, we obtain a new search window, which is highlighted by the rectangle of dotted lines

2.2 Non-dominant Color Extraction and Edge Detection

After deciding the region of interest, we can start to extract the player figure. Major detection methods in literatures include dominant color detection [2, 5, 10] and background subtraction [9]. Due to the reason that background subtraction



(a) (b)

Fig. 3. (a) Possible player locations in real court and (b) Possible player locations in image space, where the circle denotes the current location, and triangles denote possible locations

cannot handle camera viewpoint change, non-dominant color detection is employed in our system. J. Han et al presented a non-dominant color detection method in RGB color space [5]. Nevertheless, they selected the average color of full court as dominant color, which may have large deviation with the color in players neighbor. In addition, colors of different parts of the court are affected by light, shadow or camera viewpoint. To get more accurate value, we can take advantage of court knowledge and use average color of the field where player belongs to. According to the model of Fig. 2, we distinguish the court into four areas: inner field of upper court, outer field of upper court, inner field of lower court, and outer field of lower court. The court is split horizontally by net line, while inner and outer fields are defined by court lines. Figure 4 demonstrates the inner and outer fields of lower court.

$$Upper\ court\ inner\ field : [z \in (h5_{down} \cap v1_{right} \cap v5_{left}) \mid y\ of\ z\ >\ P_c]$$
$$Upper\ court\ outer\ field : [z \in h4_{up} \mid z \notin (h4_{up} \cap h5_{down} \cap v1_{right} \cap v5_{left})]$$
$$Lower\ court\ inner\ field : [z \in (h1_{up} \cap v1_{left} \cap v5_{right}) \mid y\ of\ z\ <\ P_c]$$
$$Lower\ court\ outer\ field : [z \in h2_{down} \mid z \notin (h2_{down} \cap h1_{up} \cap v1_{right} \cap v5_{left})]$$

In contrast to RGB color space [5], we select hue and value channels from HSV color space to detect non-dominant color pixels. First we calculate the mean μ and variance σ^2 of each channel in the selected region of court, then use (1) to determine the non-dominant color pixels. Parameter α is an adjustable parameter which is varied with the court conditions such as different courts or different lights of the same court. The experimental results indicate that the novel approach is robust against the varying court colors, which is demonstrated by the examples shown in Fig. 4.

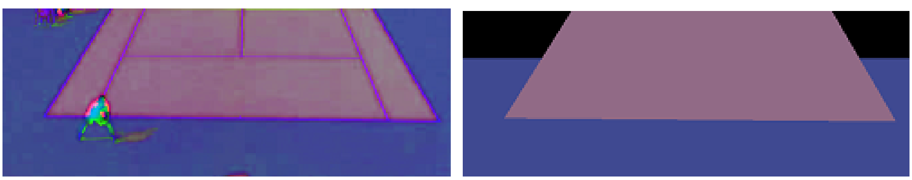
$$NDC(x,y) = \begin{cases} 1, if\|P_H - \mu_H\| > \alpha\sigma_H^2 or\|P_V - \mu_V\| > \alpha\sigma_V^2 \\ 0, otherwise \end{cases} \quad (1)$$

$$\alpha = \frac{0.5 * \beta + \sigma_H^2}{\sigma_H^2}$$

where P_H and P_V denote a pixel value of H and V channels, respectively. The β is the quantization step size of H channel. Here we quantize H into 6 dominant colors, so β is 0.5 * 1/6 = 1/12.

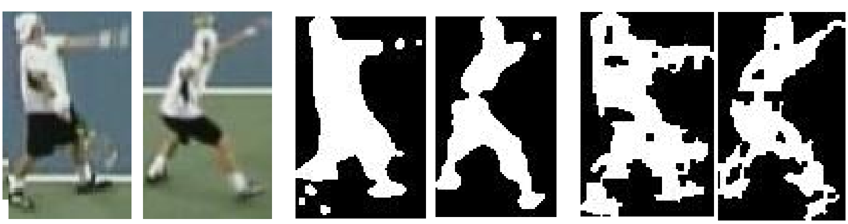
During the fierce competition of a game, players perform various actions, such as swing or serve, which may cause false detection between player body and background. In order to enhance detection reliability, we add edge detection and utilize the result to compensate non-dominant color detection. Two examples are shown in Fig. 5, where images in second column are non-dominant color extraction results, and in third column are edge detection results. As we can see, some parts of player body are lost in non-dominant color extraction but preserved in edge map, and vice versa. A well-designed combination method is capable of producing a correct and complete player body.

In terms of the edge detection flow, the data are processed by Sobel filter and generate horizontal and vertical edge images. Each image is binarized by using $\mu \pm \sigma$ as threshold. Since smoothing causes edge expansion, we need to do 1/2



(a) (b)

Fig. 4. (a) H channel image of lower court, (b) Dominant colors of inner field (purple) and outer field (blue)



(a) (b) (c)

Fig. 5. Non-dominant color extraction and edge detection results. (a) Original image, (b) Non-dominant color extraction result, (c) Edge detection result.

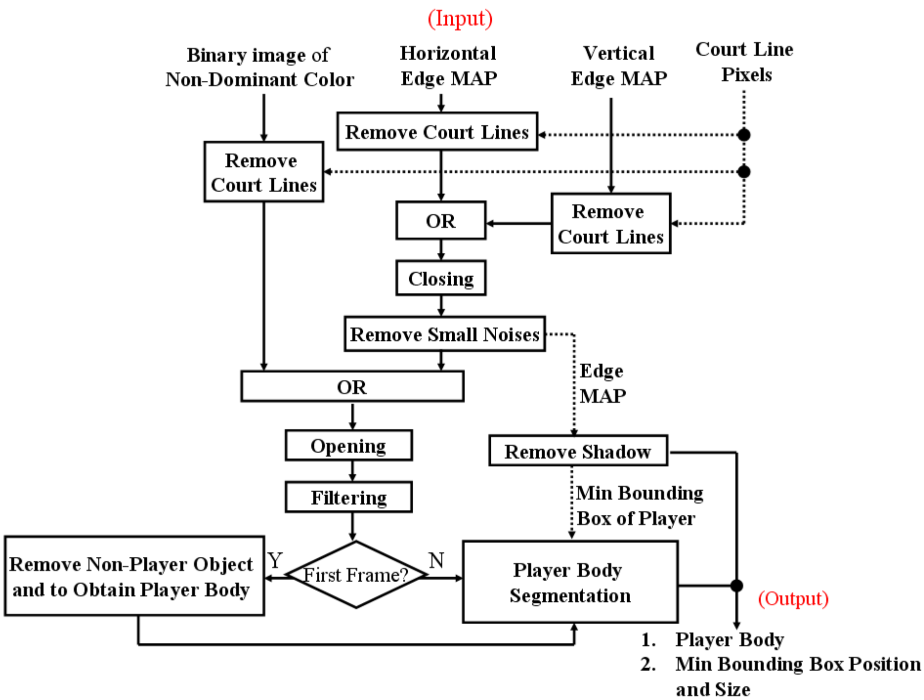


Fig. 6. Flowchart of player refinement

Fig. 7. Example of player shadows removal

sub-sampling to decrease edge width. Eventually we will get two output results, horizontal edge map and vertical edge map.

2.3 Refinement of Player Figure

The final step, refinement, is to remove undesired information and refine the player body. Figure 6 shows the flowchart of the refinement algorithm. The major steps include:

1. Remove court lines
2. Combine horizontal and vertical edge maps
3. Combine non-dominant color image and the new edge map
4. Remove cast shadow

At first, court lines must be removed from the three images. The previous developed work [12] is used to execute the job effectively. At second, we combine horizontal and vertical edge maps by performing OR and Closing operation of morphology, and use label connected components to remove noises. At third, the binary image of non-dominant color and the new edge map are merged by OR operation.

The merged result may contain shadows, thus we propose a new shadow removal technique. The shadows can be roughly classified into self shadows and cast shadows [13]. Removing self shadow is error prone and frequently eliminates parts of player body as well. Since our goal is to maintain the integrity of player figure, we concentrate on dealing with cast shadows. The color of shadow is gray or black, which has high saturation (S), and low value (V) in HSV color space. In addition, the hue (H) value is greater than that of the court color. We apply the following formulas to the edge pixels (corresponding to edge map), and the result is subtracted from the edge map, then we obtain the player figure without shadow, as illustrated in Fig. 7.

$$\begin{array}{l} -\alpha\sigma_H^2 \leq p_H - \mu_H < \frac{1}{6} \\ p_S - \mu_S \geq -\alpha\sigma_S^2 \\ p_V - \mu_V \leq -\alpha\sigma_V^2 \end{array} \tag{2}$$

3 Experimental Result

In this section, we provide experimental results of adaptive search window, player trajectory and player segmentation, respectively. The experimental data are selected from 12 videos of US Open, Wimbledon Open and French Open. The proposed algorithms are proved being robust and effective in different courts and under varying light conditions.

3.1 Adaptive Search Window

Figure 8 shows the adaptive search window (marked in black) and player window(marked in red) during the tracking period. It can be seen that both windows are changing frame by frame, which is adaptive according to the deformable player.

The results of other search algorithms [5, 8, 10] are shown in Fig. 9. The search windows are either too small or too large. Although those extraction results are sufficient for most tracking applications, they are not qualified for high level applications. Smaller search windows lead to lose parts of player body, and incur false judgment of player actions. Lager search windows contain too much noise and redundant information, make the tracking process inefficient. Due to the reasons above, the proposed adaptive window method is more suitable for high level automatic annotation system.

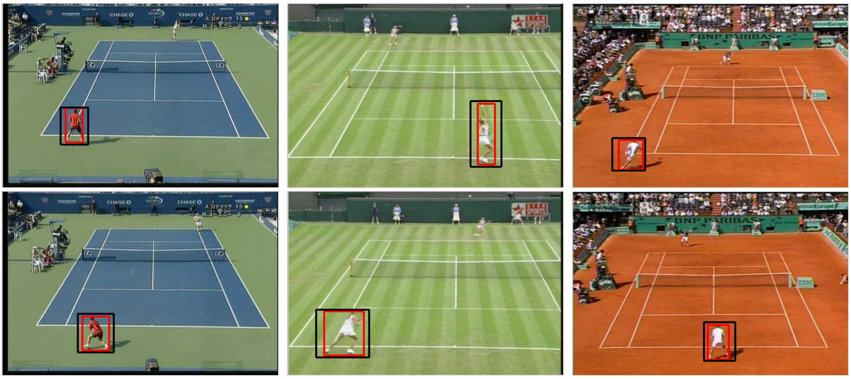


Fig. 8. Experimental results of proposed adaptive search window

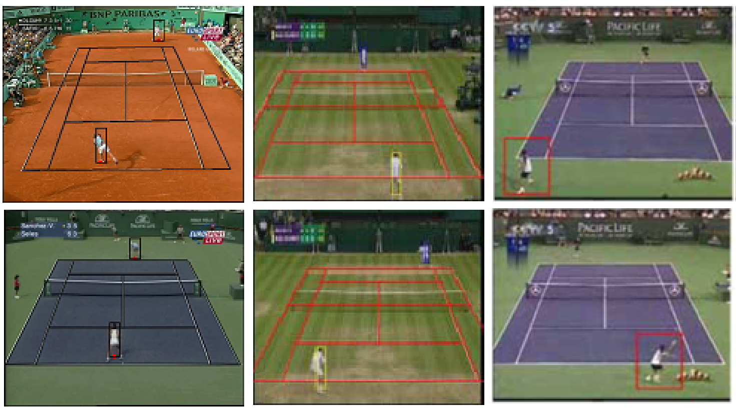


Fig. 9. Search window of other algorithms [5, 8, 10]

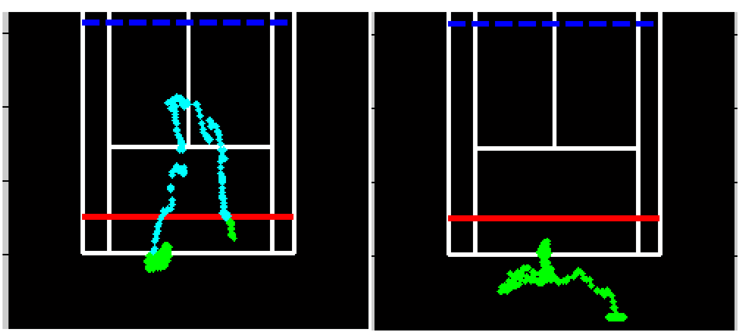


Fig. 10. Player trajectory of approach volley and ground stroke event

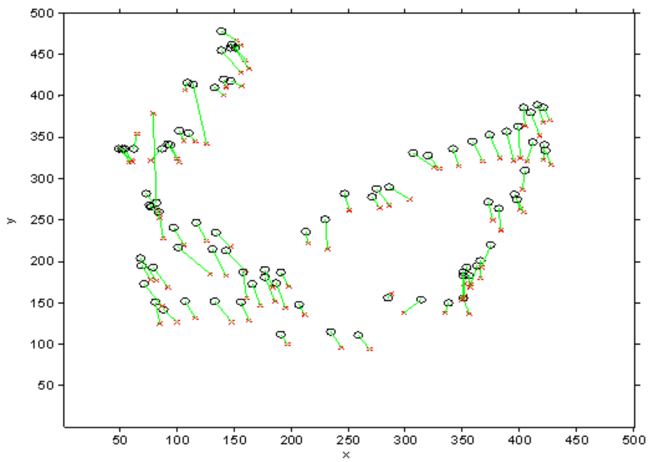


Fig. 11. o: manual tracking result, x: automatic tracking result, the connection line shows the difference between manual and automatic results

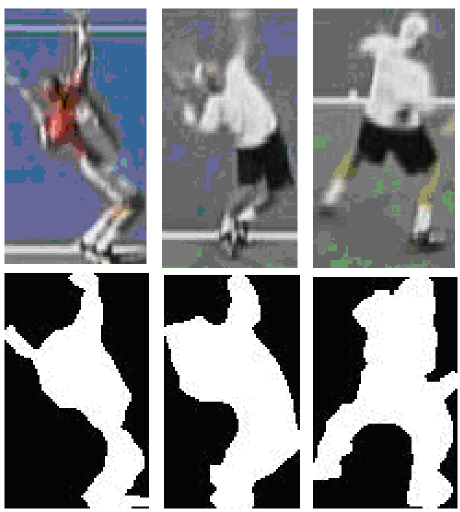


Fig. 12. Player figure extraction results

3.2 Player Trajectory

In terms of tracking players, we need to find the centroid of refined player figure for representing player position. The tracking results are shown in Fig. 10. The historic movement of approach volley event is in first image, while movement of event ground stroke is in the second. The comparison of manual and automatic tracking is shown in Fig. 11. Generally the automatic tracking results are close to manual results. Nevertheless, there are still a few mismatch errors. It is because that players white clothes, sometimes mixing with court lines, are incorrectly removed and lead to misjudge of players centroid.

3.3 Player Segmentation

The factors affecting accuracy of player segmentation are player window and segmentation algorithm. The proposed adaptive window and player body extraction algorithm are robust and effective, so we achieve excellent segmentation results, as shown in Fig. 12.

4 Conclusions

In this paper, a detection and tracking algorithm focusing on complete player figure extraction is proposed. Three schemes including adaptive search window, non-dominant color extraction filter, and edge detection filter, are developed to overcome problems of deformable player figure, various light conditions, camera viewpoint change, and different tennis courts. A novel shadow removal method is also presented to refine the player figure. Regarding with adaptive search window, we employ court knowledge and using perspective transform to calculate the search window; for non-dominant color extraction, hue and value are used as parameters and the region of interest is deliberate selected; for edge detection, a Sobel filter is applied for retrieving horizontal and vertical edge maps, which are associated with non-dominant color extraction result to refine the player figure. Around 50 video segments from 12 tennis games are used to test the algorithm. Experimental results demonstrate the new approach achieves robust player figure extraction as well as accurate movement tracking.

References

1. Zhong, D., Chang, S.-F.: Long-term moving object segmentation and tracking using spatiotemporal consistency. In: IEEE International Conference on Image Processing, Thessaloniki, Greece (October 2001)
2. Zhong, D., Chang, S.-F.: Real-time view recognition and event detection for sports video. Journal of Visual Communication and Image Representation 15, 330–347 (2004)
3. Miyamori, H., Iisaku, S.I.: Video Annotation for Content-based Retrieval using Human Behavior Analysis and Domain Knowledge. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 320–325 (2000)
4. Huang, C.-L., Shih, H.-C., Chao, C.-Y.: Semantic Analysis of Soccer Video using Dynamic Bayesian Network. IEEE Trans. on Multimedia 8(4), 749–760 (2006)
5. Han, J., Farin, D., de With, P.H.N.: Multi-level analysis of Sports Video Sequences. In: SPIE Conference on Multimedia Content Analysis, Management, and Retrieval, San Jose, USA, vol. 1 (January 2006)
6. Han, J., de With, P.H.N.: A unified and Efficient framework for Court-Net Sports Videos Analysis Using 3-D Camera Modeling. In: SPIE Electronic Imaging, San Jose, USA, vol. 1, pp. 6506–6515 (January 2007)

7. Bertini, M., Cucchiara, R., Del Bimbo, A., Prati, A.: Semantic Adaptation of Sports Video with User-centred Performance Analysis. IEEE Transactions on Multimedia 8(3), 433–443 (2006)
8. Rea, N., Dahyot, R., Kokaram, A.: Classification and Representation of Semantic Content in Broadcast Tennis Videos. In: IEEE International Conference on Image Processing, 11-14, September, 2005, vol. 3, pp. III:1204–III:1207 (2005)
9. Zivkovic, Z., Petkovic, M., van Mierlo, R.J., van Keulen, M., van der Heijden, F., Jonker, W., Rijnierse, E.: Two Video Analysis Applications Using Foreground/Background Segmentation. In: Proceedings of the VIE- 2003 Conference on Visual Information Engineering, Surrey, Guildford, pp. 310–313 (July 2003)
10. Zhu, G., Huang, Q., Xu, C., Xing, L., Gao, W., Yao, H.: Human Behavior Analysis for Highlight Ranking in Broadcast Racket Sports Video. IEEE Transactions on Multimedia 09(06), 1167–1182 (2007)
11. Sudhir, G., Lee, J.C.M., Jain, A.K.: Automatic classification of tennis video for high-level content-based retrieval. In: Proc. Int. Workshop on Content-Based Access of Image and Video Databases, Bombay, pp. 81–90 (1998)
12. Jiang, Y.C., Hsieh, C.H., Kuo, C.M., Hung, M.H.: Court Line Detection and Reconstruction for Broadcast Tennis Videos. In: IPPR Conference on Computer Vision, Graphics and Image Processing (2008)
13. Andrea Prati, I.M., Mohan, M.T., Rita, C.: Detecting Moving Shadows: Formulation, Algorithms and Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(7) (July 2003)