Tracking Racket Face in Tennis Serve Motion Using High-Speed Multiple Cameras

Hiroshi OHYA Hideo SAITO
Keio University, Japan
ohya-h@ozawa.ics.keio.ac.jp, saito@ozawa.ics.keio.ac.jp

***Abstract*— Motion capture systems have recently been used in entertainment field, for example, computer graphics, game, movie, and so on. However, it's difficult to measure motion in a natural state in existing motion capture systems, because it's necessary to wear particular equipments such as a lot of markers. This becomes a serious problem in sport motion analysis in actual play. In this paper, we propose a method for marker less motion capture in tennis play by using vision-based techniques. In this method, we track tennis racket face from motion image sequences captured with multiple cameras. We assume that the face of the racket can be modeled as an ellipse, so that we can achieve the tracking by detecting ellipse shape in the multiple view images. We employ high-speed camera system to track the racket because tennis serve motion is too high speed to track it with normal cameras. The multiple view cameras are related with fundamental matrices among the cameras, so that the position of undetected racket in one camera can be found in other positions in other cameras. This enables tracking with good accuracy of racket face. For demonstrating the efficacy of the proposed method, we capture some scene of tennis serve play in a room with multiple cameras, and then the racket can be successfully tracked with the proposed method.**

I. Introduction

In recent years, researches for sport scene analysis are very popular. We can categorize such researches for sport scene analysis into two groups. One of them is related to generation of new images for entertainment such as TV broadcasting [1]. Another is related to acquisition of data which describe features of sport player's motion, or, motion analysis. Using motion capture technology of the second category, we can acquire various data of features of player's motion.

It's common to use markers in motion analysis[2]. That is because markers are easy to be detected by sensors, so that complicated motion can easily been tracked. However, it requires hassle and a long time to wear a lot of markers. There are several kinds of method in motion capture, but we must wear particular equipments or many markers[3]. Therefore it's difficult to measure sport player's motion in a natural playing state with the conventional motion capture systems. If we can analyze sport player's motion using only images without any attachment to the players, we can capture the motion of the real playing of sport.

For motion capturing with images, the human motion must be tracked in the motion image sequence. There are many researches to track human motion by using image data. Some researchers track motion of human body or hands by using simple models representing the body rather than handling complicated shape of the body [4][5]. Other researchers track human moving indoors by using kalman filter or HMM[6][7][8].

It's very difficult to acquire detailed data of sport motion, since players move very fast in a complicated manner. Even though such players' motion seems to be difficult to be captured, instruments of sports, such as rackets, bats, clubs, etc., can be tracked relatively easier than players body. Even such instruments provide motion information for analysis of sport scene.

In this paper, we propose a new method for tracking tennis racket face from multiple view images of tennis serve scene. Tracking tennis racket provides a sort of detailed data of tennis serve motion. There are some researches to track or analyze motion in tennis serve. For example, analysis of tennis serve motion using player's color texture pattern has been proposed[9]. This method can recognize a type of the player's motion, but detailed motion data can not be obtained.

Since tennis serve motion is very fast, and size and aspect of tennis racket face against one camera rapidly vary in every frame, we use high-speed multiple cameras for tracking the tennis racket. We describe that system environment in detail in section Ⅱ.

Shape of tennis racket is similar to ellipse. We approximate tennis racket face by an ellipse for detecting and tracking it. There are several ways to track ellipse[10][11][12]. These conventional methods assume the invariability of the size and aspect of ellipse against a camera. Since size and aspect of tennis racket against a camera vary considerably, we can not use these method to track tennis racket face in tennis serve motion. We evaluate degree of overlapping between racket edge and that ellipse. We describe the way to track tennis racket face using information from a camera in section Ⅲ A.

However, this method has limitations to track racket face accurately with only a camera. We utilize fundamental matrix for using information from multiple cameras effectively. With accurate fundamental matrix, it enables to recognize complicated hand gesture[13].

We describe the way use of information from multiple cameras in section Ⅲ B. Then we describe experiments to track tennis racket face using our method and those results in section Ⅳ. At last we describe discussion about them and conclusion in section V and section VI, respectively.

II. System Environment

In this section, we describe about system environment of our method. System environment is shown in Fig. 1. We locate four high-speed cameras (200 frame per second) in doors.

Since tennis serve motion is too fast to analyze it using images taken with normal speed camera (30 frame per second), we use high-speed camera.

We cover the background with homogenous colored clothes. In terms of location of cameras, we take account into the geometry of an object player, a ball and a racket, so that they will fit into image size of every camera. If they run off the edge of image, it becomes more difficult to track them. In addition to that geometry, we also take account into epipolar geometry among cameras, so that the epipoplar lines between neighboring two cameras are not approximately parallel. Unless this condition is fulfilled, two epipolar lines do not intersect at a point but on a line, position inference with fundamental matrix is not accurate.

We assign camera number as shown in Fig. 1. We call each camera by this number in the following description.

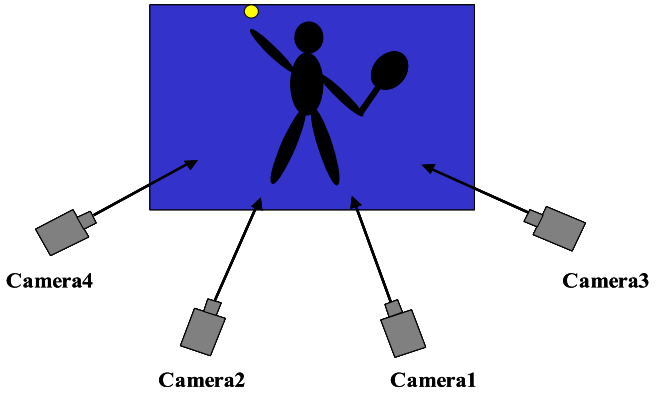


Fig. 1. System environment.

III. Tracking Racket Face

In this section, we describe about the method to track tennis racket face. Our method is categorized as inner-camera operation and inter-camera operation. We explain about each them in detail. Fig. 2 shows flow of tracking tennis racket face.

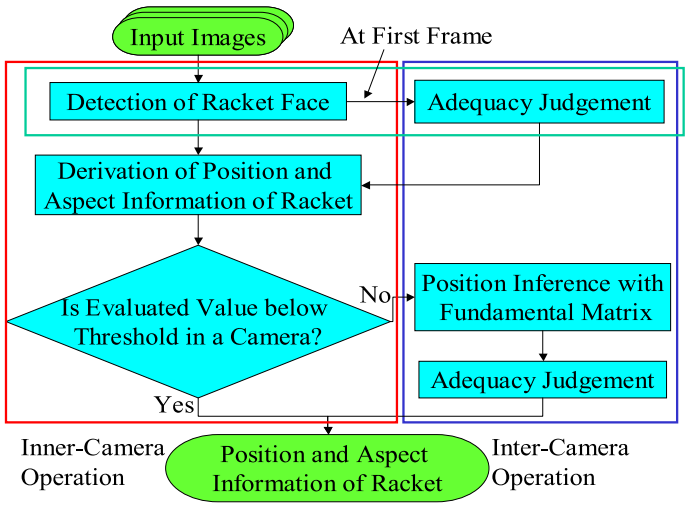


Fig. 2. Flow of tracking tennis racket face

A. Inner-Camera Operation

In this subsection, we describe about operation for images taken from a camera. When racket face tracking is missed in a camera, we estimate the position of racket face center using information obtained from other cameras. If accuracy of tracking racket face is too low in the other cameras, the estimated position is not accurate. Therefore accuracy of tracking in a camera must be better than a standard level. We give an explanation the way of tracking racket face to fulfill this condition in this section.

1) Preparation

We need extract tennis racket from input image to detect or track the racket face. Frame subtraction is done on input image to extract dynamic regions. In tennis serve motion, since the movement of racket is more definite than that of player and so on between continuous frames, extraction of racket is easy comparatively. Since we use high-speed camera, extracted dynamic regions are regarded as edges. When occlusion occurs between player and racket, however, it's very hard to discriminate between racket edge and edges of player's body. We need contrive ways to detect or track racket face in that situation.

2) Detection of the Racket Face at the First Frame

Performance function for detection of racket face is composed of three factors as mentioned below.

(a) Degree of Overlapping : As mentioned in section I, we approximate the racket face by the ellipse to track racket face. Five parameters, or x-coordinate and y-coordinate of center, major axis, minor axis, orientation of the ellipse are needed to draw an ellipse. To put it the other way around, if we give those five parameters, an ellipse is drawn in images. We can gain degree of overlapping by superposing this drawn ellipse with the edges in images. Degree of overlapping is

$$edge(x, y, lr, sr, \theta) = \frac{1}{M}(\sum_{m=0}^{M-1} white(m)), \quad (1)$$

where, x, y, lr, sr, θ are five parameters, M is the number of pixel on the drawn ellipse. $white()$ is a function that if lightness of pixel is 255, return 1, if lightness is 0, return 0. $edge(x, y, lr, sr, \theta)$ is a function that represents what proportion the pixel whose lightness is 255 exists on the ellipse drawn by five parameters. In the case that this degree of overlapping is high, we regard that edge as the racket edge.

(b) Proportion of Pixel on the Edge inside Drawn Ellipse : Looking at the racket edge in an image, there are few edges inside racket frame (Catgut can not be seen significantly in tennis serve motion). Utilizing that feature, we try to raise accuracy of tracking racket face. Proportion of pixel on the edge inside drawn ellipse is

$$inside(x, y, lr, sr, \theta) = \frac{1}{N}(\sum_{n=0}^{N-1} white(n)), \quad (2)$$

where, N is the number of pixel inside drawn ellipse.

(c) Preservation of Drawn Ellipse Size : If we track racket face by using performance function composed of two functions as stated above, major axis and minor axis of the ellipse tend to become reduced in size frame by frame. So we need a function to preserve drawn ellipse size. This function represents

$$rad(lr, sr) = \frac{1}{5}((lr - midlr) + (sr - midsr)), \quad (3)$$

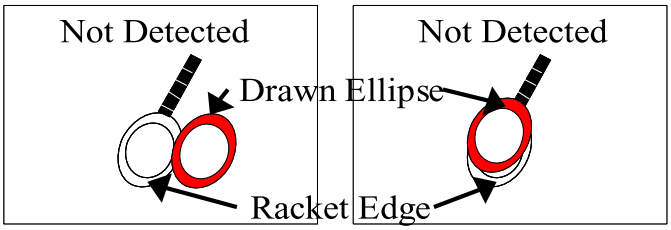
where, $midlr, midsr$ are lr, sr at previous frame, so that higher lr, sr can make higher $rad(lr, sr)$.

We detect or track tennis racket face with $eval(x, y, lr, sr, \theta)$ combined three functions as stated above. $eval(x, y, lr, sr, \theta)$ is

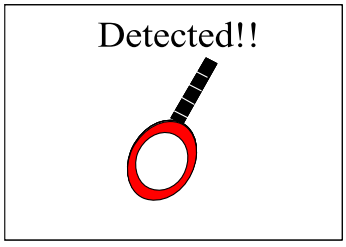
$$eval(x, y, lr, sr, \theta) = edge(x, y, lr, sr, theta) - inside(x, y, lr, sr, theta) + rad(lr, sr) \quad (4)$$

In the case that evaluated value of this function is max, we regard that edge as the racket edge.

However it takes enormous time to calculate five parameters that satisfy this condition in whole images. Therefore, we do not use an ellipse but a circle to detect approximate region of the racket face at first frame, since a circle can be specified by giving only three parameters, which are x-coordinate and y-coordinate of center and radius of the circle, so that we can decrease computation cost considerably. Scheme of detecting racket face is shown in Fig. 3.



Degree of overlapping is about (a) 10% (b) 50%



(c) Degree of overlapping is about 100%

Fig. 3. Scheme of detecting racket face

The racket face is not always detected rightly in all cameras. We check adequacy of position of detected racket face in each camera with constrained condition given by fundamental matrix. We describe about this method in detail in Ⅲ B 2).

If approximate region of the racket face is detected rightly in all cameras, the next operation is search more accurate region of racket face by using an ellipse. Since approximate region of the ellipse is already known, we can narrow down search range for five parameters. Thus it does not take long time to detect with the ellipse. We use the same performance function as the case of detecting the circle. When degree of overlapping is max, we regard that edge as accurate racket edge.

3) Tracking Racket Face

At the first frame, the racket face is detected accurately, we must track it from the second frame. The method of tracking is equal with that of detecting basically. Since we use high-speed cameras, five parameters representing the ellipse do not change very much compared to previous frame. Therefore we can narrow down search range about five parameters considerably after the first frame. So the racket face at each frame is tracked by finding optimum five parameters of the ellipse.

The result of tracking in each camera is not always accurate. If evaluation value is below the threshold which is obtained by previous experiment, we do not trust the result of tracking in the camera and correct position information of the racket face center by using the racket position detected in the other camera via fundamental matrix. In this paper, we call this situation as 'miss-tracking'. In contrast, we call the situation that evaluation value is above the threshold 'success-tracking'. We call this threshold as 'threshold1'.

Detail of this way to correct it is described in Ⅲ B 1). Then we search optimum five parameters again with a focus on corrected position of the racket face center in more narrow range than before correction about x-coordinate and y-coordinate.

4) Use of Position Information of Ball at Impact

At the frame that the racket face indicates precisely, we can track it by using the method described in Ⅲ A 3). Since swing speed is so fast at the impact frame that the racket edge is blurred or thinner, tracking racket face becomes very hard. For avoiding this problem, we use position information of a tennis ball position. Since the racket is hitting the ball, the face of the racket will meet with the ball in the final moment of the swing. This means that the distance between the racket face and the ball should be almost decreased in the swing. Thus we give the stronger weight to the term of $eval(x, y, lr, sr, \theta)$ for smaller distance between the racket face center and the position of the ball at impact. Then, $eval$ is extended to $eval2(x, y, lr, sr, \theta)$ as follows.

$$eval2(x, y, lr, sr, \theta) = eval(x, y, lr, sr, \theta) - ball(x, y) \quad (5)$$

$$ball(x, y) = |impactx - x| + |impacty - y| \quad (6)$$

where $impactx, impacty$ are x-coordinate and y-coordinate of the ball at impact which is tracked beforehand, $eval2(x, y, lr, sr, \theta)$ is performance function for tracking racket face from the frame right before impact to impact frame. We can track racket face without missing in the situation that swing is high speed by using this $eval2(x, y, lr, sr, \theta)$ for performance function.

We need to define the frame that performance function switches from $eval$ to $eval2$. We define it the frame that evaluated values of $eval(x, y, lr, sr, \theta)$ are below threshold1

in three cameras of four, after 20 frame earlier than the impact frame.

B. Inter-Camera Operation

We describe about inner-camera operation of tracking racket face in preceding subsection. In this operation, the tracking results are not always accurate, because miss-tracking sometimes occurs due to three factors as mentioned below.

(a) Being high speed of tennis serve motion. This results in getting thinner or blurring the racket edge, in consequence miss-tracking occurs.
(b) Change of aspect and appearance of the racket face. It's difficult to approximate the racket face by the ellipse and the racket face itself can not be seen due to this.
(c) Occurring occlusion. Occlusion occurs between the player itself and the racket principally, miss-tracking occurs owing to identifying player's body edge as the racket edge.

If evaluated value in one or two cameras are lower than the threshold1, we need correct the tracking results in them by using the tracking results in the other cameras via fundamental matrix. In this section, we describe about the way of that correction in detail.

1) Position Estimation with Fundamental Matrix

We correct the position in one or two miss-tracking cameras by using constrained condition given by fundamental matrix. First, we account for constrained condition given by fundamental matrix.

Let us assume that the results of detecting racket face in camera1 and camera2 are reliable, but the detected position in camera3 is not reliable. In this situation, if we draw epipolar lines corresponding to racket face center in camera1 and camera2 in image of camera3, these lines intersect at one point. This point indicates the racket face center in camera3.

If miss-tracking also occurs in camera4, the way to correct is similar to camera3. In this way, we can estimate the position of the racket in a miss-tracking camera from the racket positions detected in other two cameras, which have higher evaluated value of tracking among cameras.

If the evaluated value tracking is lower than threshold1 in a camera once, we decide not to trust the result of tracking in the camera for 30 frames. However, if the evaluated value is higher than a particular threshold (threshold2)in this 30 frames, we decide to trust in result of tracking in that camera again.

In this way, we can correct the result of miss-tracking by using position information of the racket face center in more reliable two cameras and constraint condition given by fundamental matrix. This way to correct is shown in Fig. 4.

2) Evaluation of Position by Fundamental Matrix after Detection

As mentioned in Ⅲ A 2), we need to evaluate adequacy of the position of detected racket face in each camera. We come up with the way for evaluating the adequacy of the position by using fundamental matrices among cameras.

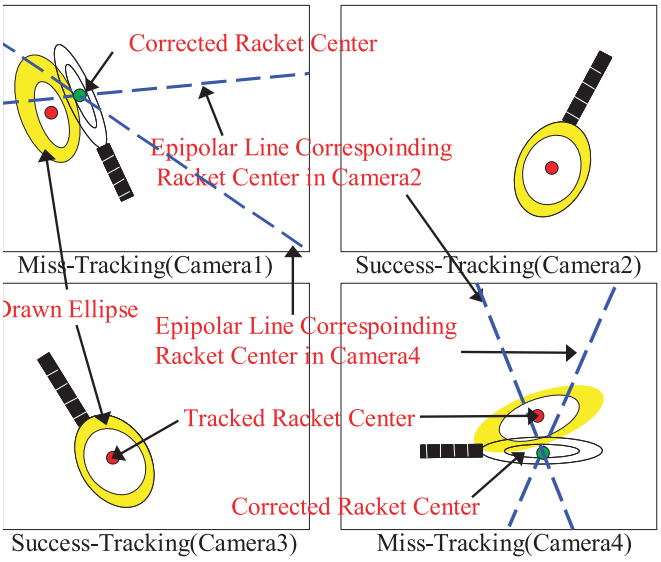


Fig. 4. Scheme of correction of tracking result

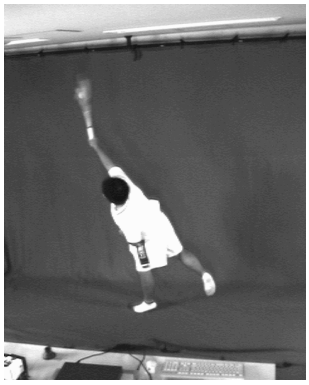
(a) Two cameras are selected among four. We assume that the results of detection in these cameras are reliable. We call these cameras as criterial cameras.
(b) We draw epipolar lines corresponding racket face center in these cameras in images of the other cameras.
(c) We compare the coordinate of intersection of drawn epipolar lines with the result of detecting racket face center in each camera. If the distance between them is close, we can consider that the results of detecting in the camera and criterial cameras are reliable.
(d) We change the combination of criterial cameras and repeat from (a) to (c) till finding three or four reliable cameras. If the number of reliable cameras is three, we correct the coordinate of the racket face center in the other camera using information of that in three reliable cameras. The way of this correction is the same as described in Ⅲ B 1).

IV. Experimental Results

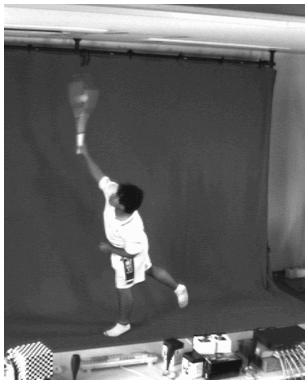
In this section, we describe experiments to demonstrate that our method for tracking the racket in tennis serve motion with multiple high-speed cameras is more effective than tracking with one camera. Input images used in this experiment are digitized of 640x480 pixels, monochrome pictures, and 200 frames/sec. The sample tennis serve scene in this experiment is composed of 123 frames. Some of input frames are shown in Fig. 5.

We will demonstrate the effectiveness of our method at the following viewpoints.

(a) detection accuracy of the racket face at the first frame. (experiment1)
(b) effect of correcting position of the racket face center with multiple cameras.(experiment2)
(c) effect which is available by using the ball position information at impact to track the racket face more precisely around impact frame.(experiment3)

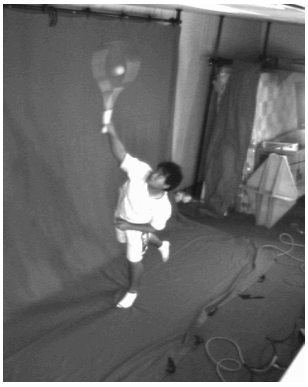


(a) Input image(Camera1)



(b) Input image(Camera2)

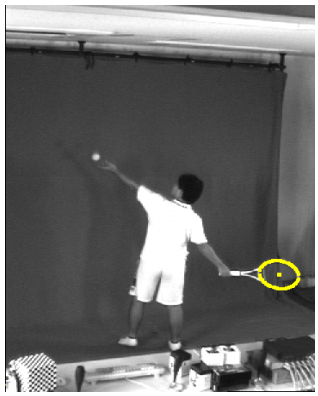
(c) Input image(Camera3)



(d) Input image(Camera4)

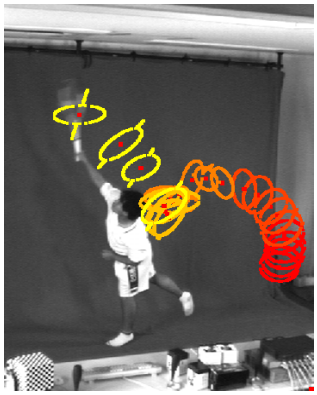
Fig. 5. Input images

(a) Using only camera2

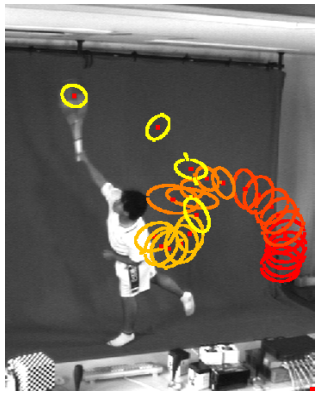


(b) Using multiple cameras

Fig. 6. Detecting tennis racket face at the first frame



(a) Using only camera2



(b) Using multiple cameras

Fig. 7. Tracked locus of the racket face

First, the result of experiment1 is shown in Fig. 6. Fig. 6 shows the result of detecting racket face at the first frame in camera2. As shown in Fig. 6 (a), if we use only image data in camera2 to detect the racket face, computer recognizes a box located at the lower left of the image as tennis racket face wrongly. This is perhaps because computer recognizes shepherd check of this box as the edge, there are a lot of edges on this box. However, if we use multiple views information, we can figure out whether the result of detecting shown in Fig. 6 (a) is wrong. Then we can correct it as shown in Fig. 6 (b).

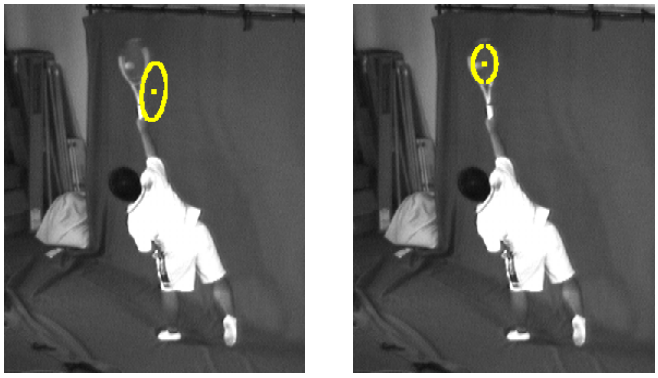
Second, result of experiment2 is shown in Fig. 7. Fig. 7 show the result of tracking racket face through image sequence in camera2. As shown in Fig. 7 (a), if we use only image data in camera2 to track the racket face, computer recognizes edges of head and shoulder of player as the edge of tennis racket face wrongly. This is perhaps because occlusion occurs between player and racket, computer can not distinguish between the racket edge and the edge of player's body. Moreover, after miss-tracking occurs because of this occlusion once, the racket face can not be tracked accurately. If we use multiple views information and fundamental matrix, however, we can track accurately as shown in Fig. 7 (b). Even if miss-tracking occurs at a frame in a camera, we can correct the result of tracking at the frame. This correction improves accuracy of tracking racket face in the camera.

Third, the result of experiment3 is shown in Fig. 8. Fig. 8 show the comparison of tracking racket face at a frame around impact in camera3 in the case of using ball position information at impact and not. As shown in Fig. 8 (a), if we do not use ball position information at impact, swing speed is too fast to track the racket face accurately. However, if we use the ball position information at impact, tracking succeed around exactly as shown in Fig. 8 (b). These two tracking are done on a like-for-like basis except a condition about ball position information at impact. Therefore it's proved that to use ball position information at impact is effective to track the racket face at around impact frame.

V. Discussion

In this section, we discuss about the tracking results. Table. I, II, III, IV show the comparison of the tracking accuracy in two cases, or the case to track using only a camera and multiple cameras, in other words, our method. According to these tables, tracking success rate is higher about from ten to forty percents by using our method than a camera alone in all cameras. This proves the effectivity of our method.

By the way, Although the frame where the whole rackets hide entirely does not exist, we miss tracking the racket face in more than three cameras about during 20 frames, except around impact. In the meantime, since three problems which



(a) Without ball position (b) With ball position

Fig. 8. Effect of the ball position for tracking result

mentioned in III B has occured in almost all cameras, we can not correct the position of the racket face center by our method. So we examine about other ways to raise tracking accuracy here. There are three ways to come to our minds.

One is changing racket color, or lightness. In this experiment, racket frame and player's wear are white, background is blue back. So we can distinguish the racket from background easily, but can not distinguish the racket frame from the player easily. Therefore if occlusion occurs between the racket and the player, miss-tracking tends to occur easily. Then if racket color changes in order to distinguish the racket from player and background, tracking accuracy may rise.

Second is the way to correct blurring. Even if we use high-speed camera(200fps), tennis serve motion is too fast, and the racket edge in images get thinner or blur. If we can correct this blurring, or if we can use a camera with higher shutter speed, miss-tracking may be decreased.

The last is making reviews on camera location. If we can capture the tennis serve motion from various view point angle, we can correct the position of the racket face center with fundamental matrix effectively so that we may raise tracking success rate. Alternatively, if we increase number of cameras, we can expect that tracking success rate may rise.

It takes approximate 40 minutes to track the racket face through 123 frames by a PC with AMD Athlon(tm) 1.73GHz. Recucing the computation cost is one of our research issues in future.

VI. Conclusion

We proposed the method that enables track tennis racket face without any markers using only images which are taken with four high-speed cameras. We constructed the system to gain position information of racket face center by applying fundamental matrix between each camera. Even if tracking of racket face is failed in one or two cameras, we can estimate approximately exact position of racket face center in them by applying this system. Therefore we can track racket face in tennis serve scene more accurately by using the proposed method.

TABLE I

Comparison of Accuracy of Tracking Racket for Camera1

	Only Camera1	Our Method
All Frames	123	123
Tracking Success Frames	64	99
Tracking Success Rate(%)	52.0	80.5

TABLE II

Comparison of Accuracy of Tracking Racket for Camera2

	Only Camera2	Our Method
All Frames	123	123
Tracking Success Frames	56	102
Tracking Success Rate(%)	45.5	82.9

TABLE III

Comparison of Accuracy of Tracking Racket for Camera3

	Only Camera3	Our Method
All Frames	123	123
Tracking Success Frames	46	107
Tracking Success Rate(%)	37.4	87.0

TABLE IV

Comparison of Accuracy of Tracking Racket for Camera4

	Only Camera4	Our Method
All Frames	123	123
Tracking Success Frames	60	76
Tracking Success Rate(%)	48.8	61.8

References

[1] N.Inamoto and H.Saito: "Intermediate View Generation of Soccer Scene from Multiple Videos", Proc. of International Conference on Pattern Recognition (ICPR 2002)

[2] Pascual J. Figueroaa, Neucimar J. Leitea and Ricardo M. L. Barros, b: "A flexible software for tracking of markers used in human motion analysis" Computer Methods and Programs in Biomedicine 72, Issue 2, pp.155-165, October 2003

[3] Hiroyuki Okamoto, Takahito Suzuki, Manabu Kuromori and Hidetsugu Ichihara: "The motion capture system" Information Processing Society of Japan, Computer Vision and Image Media, 128-13, pp97-102, 2001

[4] Quentin Delamarre, Olivier Faugeras: "3D Articulated Models and Multi-View Tracking with Silhouettes", IEEE International Conference on Computer Vision September, pp.20-27,1999

[5] G.McAllister, S.J.McKenna, I.W.Ricketts: "Hand tracking for behaviour understanding", Image and Vision Computing 20, pp.827-840, 2002

[6] Dae-Sik Jang, Seok-Woo Jang, Hyung-Il Choi: "2D human body tracking with Structural Kalman filter", Pattern Recognition 35,pp2041-2049,2002

[7] I.A.Karaulova, P.M.Hall, A.D.Marshall: "Tracking people in three dimensions using a hierarchical model of dynamics", Image and Vision Computing 20,pp691-700,2002

[8] Alexandra Psarrou, Shaogang Gang, Michael Walter: "Recognition of human gestures and behaviour based on motion trajectories", Image and Vision Computing 20,pp349-358,2002

[9] Keiko Yoshinari, Ryoso Tomosue, Hiromi Okazaki, Harumi Iwasaki, Yoshifuru Saito and Kiyoshi Horii: "Proper Tennis Serve Analysis Using Cognition Method of Eigen Pattern Image" Journal of The Visualization Society of Japan, 23, Suppl.No.1, pp279-282, 2003

[10] Sung Joon Ahn,Wolfgang Rauh, Hans-Jürgen Warnecke: "Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola", Pattern Recognition 34,pp2283-2303,2001

[11] Xiaoming Zhang, Paul L.Rosin: "Superellipse fitting to partial data", Pattern Recognition 36,pp743-752,2003

[12] Mitsuo Okabe, kenichi Kanatani and Naoya Ohta: "Automatic Ellipse Fitting by Ellipse Growing Information Processing Society of Japan, Computer Vision and Image Media, 126-002, pp9-16, 2001

[13] Xiaoming Yin , Ming Xie: "Estimation of the fundamental matrix from uncalibrated stereo hand images for 3D hand gesture recognition", Pattern Recognition 36, pp.567-584, 2003