Proceedings of 2017 IEEE
International Conference on Mechatronics and Automation
August 6 - 9, Takamatsu, Japan

Fusion of Vision and IMU to Track the Racket Trajectory in Real Time

Kun Zhang, Zaojun Fang, Jianran Liu, Zhengxing Wu, Min Tan
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences, Beijing,China
zhangkun2012, zaojun.fang, liujianran2012, zhengxing.wu, min.tan @ia.ac.cn

***Abstract*— This paper presents a novel approach on identification of human habits and behaviors when playing with the table tennis robot. The main idea of this approach lies in tracking the trajectory of the racket on player's hand by fusion of vision and IMU(Inertial measurement unit) sensors' data. In order to uniform the data from the vision and IMU sensors, the non-linear algorithm is applied to accomplish the calibration. The visual viewable range could be broadened by the method to switch the vision systems between the monocular and binocular vision system. The fusion approach of the vision and IMU sensors is based on the EKF (Extended Kalman Filter) for obtaining accurate and robust racket pose. Taking advantage of the racket pose, the player's habits and behaviors can be represented when he playing with the table tennis robot. Experiments and results showed that the proposed method is effective and real-time.**

***Index Terms*— Racket pose, vision and IMU sensors, non-linear method, Extended Kalman Filter**

I. Introduction

Using robots to play balls is no longer a novel idea, such as the table tennis robots, the badminton robots, etc. However, they could not complete the games like human beings perfectly. The main reason is that the robot is lack of intelligence and flexiblity. As a representative of the robot playing ball, the table tennis robot is a complex system, involving the automatic control, machine vision and trajectory prediction algorithms.

If a table tennis robot wants to play well, it must have two indispensable parts. One is the mechanism to hit the ball back to play the part of our arms. The other is a visual system to obtain ball positions accurately as its eyes. Therefore, optimization of a table tennis robot is mainly concentrated on the two parts. In the early period, the arms of table tennis robot were mainly industrial robotic arms, for example, the robot arm with 7 degrees of freedom (DOF) from Toshiba [1], the 6-DOF PUMA 260 chosen by Andersson [2]. In last few years, researchers were more willing to use diverse mechanisms to organize a table tennis robot arm instead of the industrial robotic arms. Acosta constructed a low-cost table tennis robot with a 2-DOF mechanism and the robot processor was based on the PC [3]. With the popularity of humanoid robots, table tennis robots were replaced as humanoid robots. The representative robots were the 'Kong' designed by ZheJiang university [4] and the 'HuiTong' produced by Beijing Institute of Technology [5].

As for table tennis robots' visual systems, they develops from complex to simple. Acosta obtained the ball position by the relationship between the light position, the camera position and the ball's shadow using a monocular vision system. To get the 3D (three-dimensional) coordinates of the ball, using the binocular vision system was a simple way.

As for the table tennis robot, knowing more information about the flying ball is good for striking back by robots. For the opponent, his action to strike the ball back is also one important information, which is helpful for the table tennis robot. Wang presented a framework of anticipatory for table tennis robot to handle human behaviors [6].

The racket pose includes its position and orientation. Chen combined perspective-n-point(PNP) and orthogonal iteration(OI) algorithms to obtain the racket pose relying on a monocular vision system [7]. However, the monocular vision system is hard to compute 3D coordinates. In addition, visual measurement is unstable because the working environment is complex and the racket pose changes fast. Measuring one object's orientation, the gyroscope is suitable if it has no gyro drift. Therefore, we choose an IMU to measure the racket orientation because an IMU basically contains gyroscopes, acceleration and other modules.

This paper focuses on obtaining the racket pose. For obtaining the racket trajectory, a novel method is proposed. In section II, the table tennis robot is described. The way to acquire the sensors data is presented in section III. In section IV, the racket pose could be computed by fusing the data from IMU and cameras. The experiments and results are shown in V. At the end, a simple conclusion is given in section VI.

II. System Description

The research platform is shown in Fig.1, it includes three parts, the motioning mechanism, a set of visual system

This work was supported by the National Natural Science Foundation (NNSF) of China under Grant 61305024, Grant 61273337, Grant 61227804 and Grant 61333016. It was also supported by the Beijing Natural Science Foundation under Grant 3141002.

978-1-5090-6759-6/17/$31.00 ©2017 IEEE

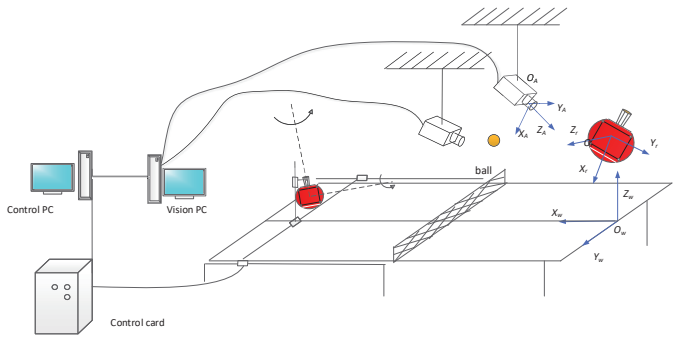


Fig. 1. The table tennis system

and the opponent's racket on hand. The motion mechanism contains a 5-DOF arm to return the ball and the control system which consists of a PC and one motion control card. The visual system has two high-speed cameras and a PC for image processing. The two cameras are fixed above the middle of the table to collect the opponent's behavior. For extracting the racket pose quickly, some marks are drawn on the racket's positive side, as shown in Fig. 2. The marks are a black rectangle with four white corners and a white line parallel to the rectangle. The IMU module is placed on the racket's negative side.

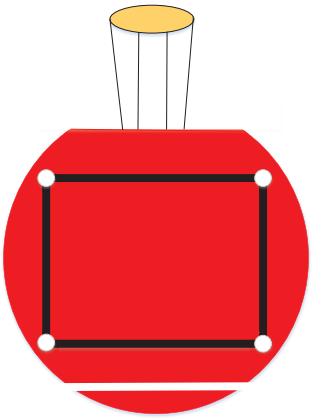


Fig. 2. The racket's mark

To describe this system, a set of coordinate frames are defined. The list is shown as follows:

- F_w:(O_w, X_w, Y_w, Z_w) is the world coordinate frame. It is situated on the table plane. The middle point of table short side, which is close to the opponent side, is chosen as the ordinate origin. Its Z_w axis is vertically upward.
- F_c:(O_c, X_c, Y_c, Z_c) is the camera coordinate frame established on the camera's optical center. The optical direction is selected as the Z_c axis, and the horizontal direction is the X_c axis.
- F_r:(O_r, X_r, Y_r, Z_r) is the racket coordinate frame. Its coordinate origin is set at the racket's center. Its Z_r axis is the racket normal direction and the direction of its X_r axis points toward the racket's handle.
- F_g:(O_g, X_g, Y_g, Z_g) is the IMU coordinate frame. Its directions of all axes are almost the same as the racket coordinate frame. However, it also need be calibrated.
- F_e:(O_e, X_e, Y_e, Z_e) is the earth coordinate frame. It is established according to the earth magnetic field.

III. The Acquisition of Sensors Data

This section discusses how to obtain racket pose from sensors. The racket pose could be computed if the corners' ordinates in the image coordinate frame are known. The IMU could only provide the racket orientation instead of the racket position.

A. Image Processing For Obtaining Racket Position and Pose

Generally, the racket positive plane is red, which is a great feature easily extracted. Then the four corners can be detected from the red region. However, During the striking process, the racket pose is fast-changing. As a result, the light on the racket is unstable. What's more, the racket positive plane would reflect the light. The racket in the images sometime becomes total black or white. Hence, we adopt the self-learning method based on characters between the HSV and RGB color space [8]. The flowchart of image processing is shown as Fig.3, it includes 6 steps.

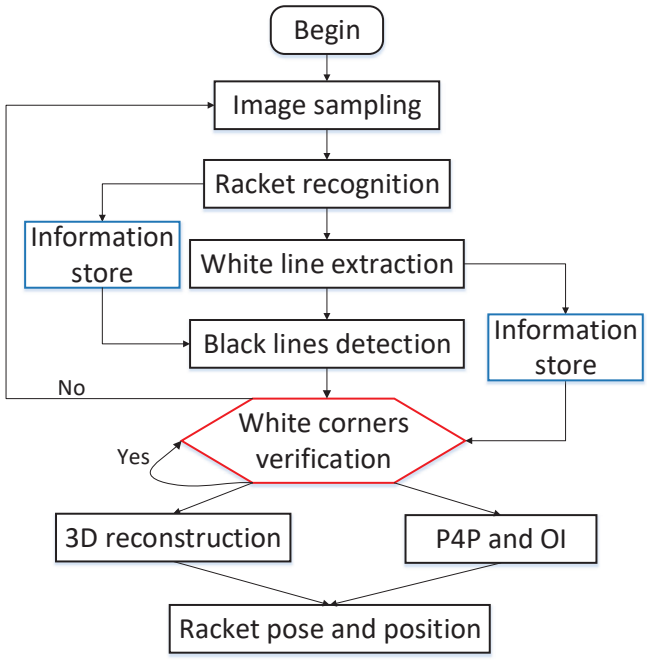


Fig. 3. The flowchart of image processing.

1) Finding the racket position in images is the first step. In the HSV color space, the H component value of

red pixel is close to 0° or 360°. If the pixel's S component value also reach certain value, the pixel can be considered as a red pixel instead of black pixel. Thus, the racket could be detected. At the same time, all the components' values in the RGB color space of the red pixels could be stored and used the next steps because those values indirectly report the light circumstance on the racket. The racket center could be acquired by counting all the red points coordinates.

2) In the HSV color space, the white pixel's V component value is obvious higher than other color pixel. Moreover, the white line is surrounded by the racket. According to the two rules, we could extract the white line and calculate the white line function. For making the white line function more accurate, the line refinement is necessary. And we also store those component values of the white points in RGB color space for the following steps.
3) The slopes of the four black lines in the rectangle are known after the white line function is gotten. The component values of black points in the RGB color space are less than the red points. With the help of the stored information from step 1, the black points could be found by scanning line by line. The least square method is applied to compute the black lines' functions. Each black line's corresponding position is distinguished from the relative position relations with the racket's center and the white line.
4) The four corners of the black rectangle could be obtained with the black lines' functions. However, the real corners may be deviated from the computed corners because of the errors from the black lines' functions. To solve this problem, we depend on the stored information from step 2 to confirm whether the corners are the centers of little white region.
5) When both of the two cameras detect the corners in their sampling images, the coordinates of four corners in the world coordinate frame could be calculated directly by 3D reconstruction method. Based on the SVD (Singular Value Decomposition) method, the racket pose could be obtained. If only one camera finds the corners, the P4P method and OI (Orthogonal Iteration) method can be used for counting the result.
6) If the corners are extracted in one frame, the corners in continuous frame can be predicted through sliding window method. The stored information in step 2 will confirm whether the predicted corners are white points to save time and ensure the real-time. When the points are not the real corners, we have to begin to search the racket from the whole image.

B. IMU Information Handling

The racket orientation could be calculated by the cameras and the IMU. The results obtained by the cameras are easy to transfer to the world coordinate frame by the calibrated parameters. However, the racket orientation read from the IMU is relative to the earth coordinate frame. Thus, it also need calibrated to transfer from the IMU coordinate frame to world coordinate frame. The transformation between different coordinate frames are shown as follows:

$$\begin{cases} F_w = {}^wR_cF_c \\ F_w = {}^wR_rF_r \\ F_w = {}^wR_eF_e \\ F_e = {}^eR_gF_g \\ F_r = {}^rR_gF_g \end{cases} \tag{1}$$

where ${}^wR_c$ is the transformation matrix from the camera coordinate frame to the world coordinate frame, ${}^wR_r$ is the transformation matrix from the racket coordinate frame to the world coordinate frame. ${}^wR_e$ is the transformation matrix from the earth coordinate frame to the world coordinate frame. ${}^eR_g$ is the transformation matrix from the IMU coordinate frame to the earth coordinate frame. ${}^rR_g$ is the transformation matrix from the racket coordinate frame to the IMU coordinate frame. F_w, F_c, F_r, F_e and F_g are the coordinates of one point in different coordinate frames. According to those corresponding relationships.

$${}^wR_rF_r = {}^wR_cF_c \tag{2}$$

$${}^wR_r{}^rR_gF_g = {}^wR_e{}^eR_gF_g \tag{3}$$

F_c can be obtained through image processing, and F_r is prior known. The homogeneous transformation matrix ${}^wR_c$ is the calibrated parameters of camera calibrations. Thus, the homogeneous transformation matrix ${}^wR_r$ can be calculated by Eq.2. The ${}^eR_g$ could be read from the IMU. The corresponding relationships between the racket and the IMU, even though their coordinate frames are almost the same. The ${}^wR_e$ is also unknown because it could obtained directly. Based on the Eq.3, we could find that the problem is a typical hand-eye calibration problem, $AX = ZB$.

There are many methods to solve this problem [9], [11], [10], most of them belong to the linear method. However, the linear method can not guarantee the quality of solution. It may obtain some result with large errors. According to Dornaika's paper [9], the non-linear method is a better way than other method. The error function is shown as follow:

$$\begin{aligned} f({}^rR_g, {}^wR_e) = \mu_1 \sum_{i=1}^{n} (\|{}^wR_{ri}{}^rR_g - {}^wR_e{}^eR_{gi}\|^2) \\ +\mu_2 \sum_{i=1}^{n} (\left\|{}^rR_g{}^rR_g^T - I\right\|^2) + \mu_3 \sum_{i=1}^{n} (\left\|{}^wR_e{}^wR_e^T - I\right\|^2) \end{aligned} \tag{4}$$

The problem is minimizing the $f({}^rR_g, {}^wR_e)$, it could be considered as a nonlinear least-squares constrained minimization problem. The Newton's method is a good way to solve this

problem. Then, the IMU data could be transformed from the IMU coordinate frame to the world coordinate frame.

IV. Fusion of The Cameras and IMU Data

To obtain an object's pose, the gyro is a good choose. However, the gyro drift is a big problem. To control the gyro drift, the gyroscope, accelerometer and magnetometer compose of a 9 DOF IMU module. Fusion of those sensor, the gyro drift could be only suppressed but not completely eliminated. Generally the result from the IMU is precise and quick if there is no gyro drift. Two cameras become a binocular vision system when the two cameras extract the corners successfully. The calculating speed of the binocular vision system is faster than the monocular vision system because the reconstruction from the monocular vision system has an iteration step. However, two monocular vision systems have broader vision range than one binocular vision system. As for vision measurement, it has more or less errors because of the errors from the corners' pixels, especially for the monocular vision system. In a word, each sensor has its advantages and disadvantages. Thus, it is necessary to fuse all the sensors' data. Considering that the relationship between the data is non-linear, a simple EKF can be applied to solve the problem of fusing all the data [12].

The racket state X is shown as following:

$$X = [P, V, Q, \boldsymbol{\omega}] \tag{5}$$

where P represents the coordinates of the racket's center $[x, y, z]$. V is the speed of the racket's center point $[v_x, v_y, v_z]$. The quaternion Q defines the racket orientation $[q_0, q_1, q_2, q_3]$, and the $\boldsymbol{\omega}$ denotes the angular velocity of the racket $[\omega_x, \omega_y, \omega_z]$.

The measured data from all the sensors includes the racket position and pose from the visual measurement and the pose and angular velocity from the IMU module. Thus, the measured vector is defined as follows:

$$Z = [P_{cw}, Q_{cw}, Q_{gw}, \boldsymbol{\omega}_{gw}] \tag{6}$$

where P_cw and Q_{cw} are defined as the racket position $[x_{cw}, y_{cw}, z_{cw}]$ and pose $[q_{0_{cw}}, q_{1_{cw}}, q_{2_{cw}}, q_{3_{cw}}]$ from the visual measurement. Q_{gw} and $\boldsymbol{\omega}_{gw}$ are the racket pose $[x_{gw}, y_{gw}, z_{gw}]$ and angular velocity $[\omega_{x_{gw}}, \omega_{y_{gw}}, \omega_{z_{gw}}]$ obtained by the IMU module. To uniform all the sensors' data, all the measured result are transformed into the world coordinate frame.

As discussed above, the EKF is designed as follows:

$$\begin{aligned} X_{i+1|i} &= FX_{i|i} \\ P_{i+1|i} &= FP_{i|i}F^T + Q_i \\ K_{i+1} &= P_{i+1|i}H_{i+1}^T(R_{i+1} + H_{i+1}P_{i+1|i}H_{i+1}^T)^{-1} \\ X_{i+1|i+1} &= X_{i+1|i} + K_{i+1}(x_i - \hat{x}_i) \\ P_{i+1|i+1} &= [I - K_{i+1}H_{i+1}]P_{i+1|i} \end{aligned} \tag{7}$$

where F and $P_{i+1|i}$ are the state transition matrix and the state covariance matrix,Q_i and R_{i+1} are the noise covariance matrix in the process and measurement. K_{i+1} are the Kalman gain. x_i and $\hat{x}_i$ denote the measured racket's state and the predicted racket's state.

To reduce computation of the control computer, the computation of the IMU module is separated. The IMU module is controlled by one DSP(Digital Signal processor). We design another EKF to deal with the data from the IMU module.

The state vector X_g of the IMU module consists of euler angles θ_g, the angular velocity $\boldsymbol{\omega}_g$ and gyro bias B_g shown as Eq.8.

$$X_g = [\theta_g, \omega_g, B_g] \tag{8}$$

where

$$\begin{aligned} \theta_g &= [\theta_{gx}, \theta_{gy}, \theta_{gz}] \\ \omega_g &= [\omega_{gx}, \omega_{gy}, \omega_{gz}] \\ B_g &= [B_{gx}, B_{gy}, B_{gz}] \end{aligned} \tag{9}$$

The measurement data include the raw gyro measurements, normalized accelerometer vector and the yaw angle calculated by the magnetometer. Based on it, we define the measured vector Z_g as follows

$$Z_g = [\tilde{\omega}_g, \tilde{a}_g, \theta_y] \tag{10}$$

As the fusion of the IMU and visual measurement, the EKF for the 9 DOF IMU has the same steps to calculate the racket orientation. Thus, the flowchart for fusion of the cameras and IMU data is presented at Fig.4.

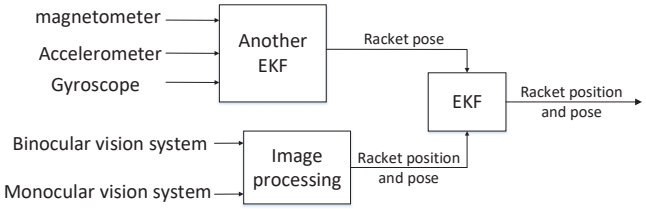


Fig. 4. The flowchart of sensors fusion.

V. Experiments and Results

A. Experiment system

To testify the proposed fusion filter, a set of experiments were conducted. The cameras were Prosilica GC660C made by Germany, whose frame rate could reach 119fps. The IMU module was placed at the racket negative plane and controlled by a STM32 chip. The IMU module communicated with the controlling computer via the bluetooth, as shown in Fig.5

B. Hand-eye Calibrating for IMU Data

First of all, We calculated the transform matrix ${}^wR_r$ by image processing and read directly ${}^eR_g$ from the IMU. As the transform matrix ${}^wR_r$ and ${}^eR_g$ were provided, ${}^rR_g$ and

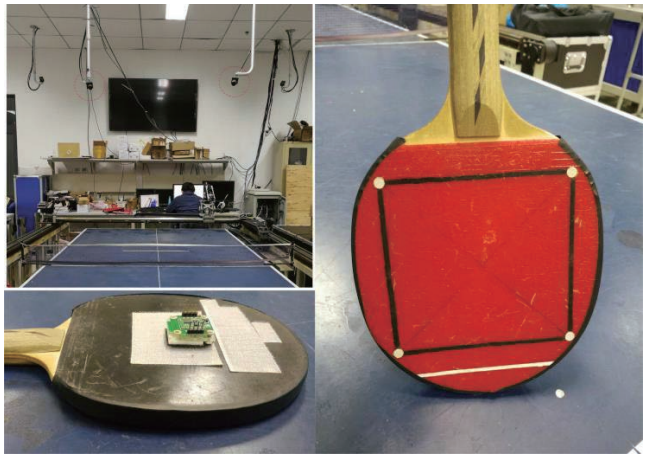


Fig. 5. The research planform.

${}^{w}R_e$ could be obtained by the proposed method. Then, the ${}^{w}R_r$ could be computed backward based on Eq.11.

$$ {}^{w}R_r = {}^{w}R_e {}^{e}R_g {}^{r}R_g {}^{-1} \tag{11} $$

Compared the racket orientation obtained by image processing, the average angle errors was shown in Table I. Obviously, the data from IMU and image processing were uniformed well using the presented method.

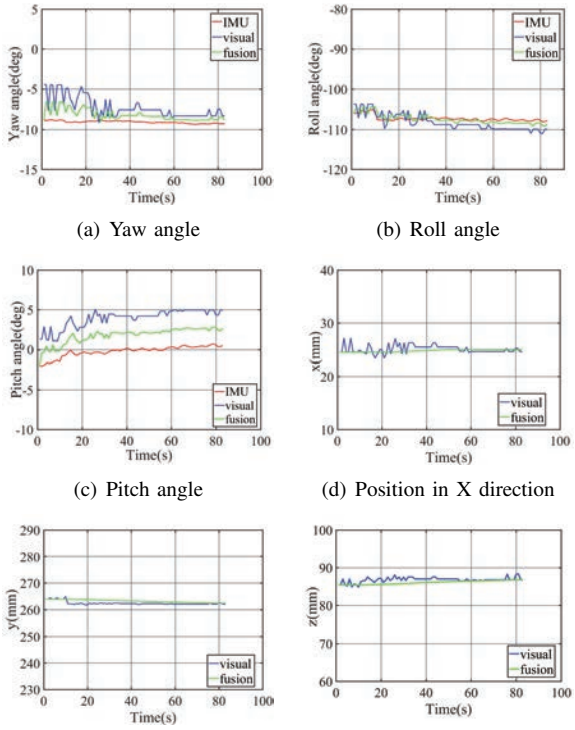
TABLE I

AVERAGE ANGLE ERRORS ABOUT RACKET POSE

	Average angle error backward calculating(deg)
Yaw angle	0.8981
Roll angle	1.6331
Pitch angle	0.7818

C. Fusion of the Data from IMU and Image Processing

In this case, we kept the racket stationary and observed the results obtained from the IMU, the visual measurement and fusion by the EKF. Fig.6 showed that the fused data, the IMU data and the result from visual measurement. It can be seen that the racket position from fused data was smoothly than the visual measurement. As for the racket orientation, The fusing method neutralized the data of visual measurement and the IMU. To test the effect when the racket stroke the flying ball, the racket trajectory was shown as Fig.7. The process included one action that racket moved forward to strike the flying ball and the test about the cameras' viewable range. The result showed that the fusion filter was working well and the cameras' viewable range was wider than only using binocular vision system due to the switch between the monocular and binocular vision system.



(a) Yaw angle (b) Roll angle

(c) Pitch angle (d) Position in X direction

(e) Position in Y direction (f) Position in Z direction

Fig. 6. The fusion results.

VI. CONCLUSION

This paper focuses on how to get the racket pose. In order to guarantee the accuracy of the racket pose, the racket pose is obtained by fusion of the visual and IMU data using the EKF. For completing the fusion of IMU and vision data, the non-linear method is used to calibrate and uniform the two sensors. And a novel method is presented that the visual system can be switched between the monocular and binocular vision system to get a broader vision and guarantees the algorithm real-time. With the racket trajectory, once the opponent completes striking, the table robot can know the hitting point immediately, it is helpful to win more time for the robot control and motion.

REFERENCES

[1] T. Broaden, "Toshiba progresses towards sensory control in realtime," The Industrial robot, vol. 14, no. 1, pp. 50-52, 1987.

[2] R.L. Anderson, "A robot ping-pong player: experiment in real-time intelligent control," MIT press, 1988.

[3] L. Acosta, J.J. Rodrigo, J.A. Mendez, G.N. Marichal, M. Sigut, "Ping-pong player prototype," IEEE robotics and automation magazine, vol. 10, no. 4, pp. 44-52, 2003.

[4] Y. Sun, R. Xiong, Q. Zhu, J. Wu, and J. Chu, Balance motion generation for a humanoid robot playing table tennis, in IEEE-RAS International Conference on Humanoid Robots, pp. 19-25, October, 2011.

[5] X. Chen, Y. Tian, Q. Huang, W. Zhang and Z. Yu "Dynamic model based ball trajectory prediction for a robot ping-pong player," IEEE International Conference on Robotics and Biomimetics(ROBIO), pp. 603-608, December, 2010.

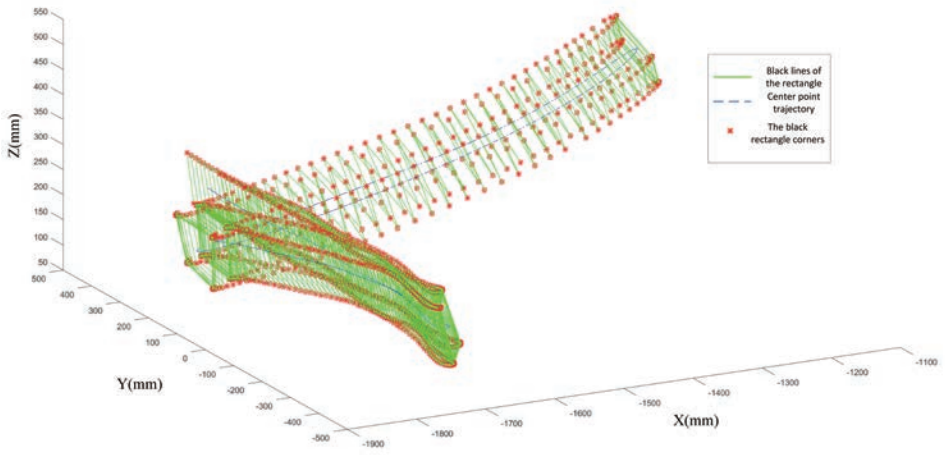


Fig. 7. The racket trajectory after fusion.

[6] Z. Wang, A. Boularias, K. Mulling, B. Scholkopf, and J. Peters, Anticipatory action selection for humanCrobot table tennis, Artificial Intelligence, 2014.

[7] G. Chen, D. Xu, Z. Fang, Z. Jiang, and M. Tan, Visual measurement of the racket trajectory in spinning ball striking for table tennis player, IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 11, pp. 2901-2911, 2013

[8] K. Zhang, Z. Fang, J. Liu and M. Tan, "An adaptive way to detect the racket of the table tennis robot based on HSV and RGB," IEEE 34th Chinese Control Conference (CCC), pp. 5936-5940, July, 2015.

[9] F. Dornaika, R. Horaud, "Simultaneous robot-world and hand-eye calibration," IEEE transactions on Robotics and Automation, vol. 14, no. 4, pp. 617-622, 1998.

[10] K. H. Strobl, G. Hirzinger, "Optimal hand-eye calibration," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4647-4653, October, 2006.

[11] R. Horaud, F. Dornaika, "Hand-eye calibration," The international journal of robotics research, vol. 14, no. 3, pp. 195-210, 1995.

[12] K. Kumar, A. Varghese, P.K. Reddy, et al. "An improved tracking using IMU and Vision fusion for Mobile Augmented Reality applications," arXiv preprint arXiv:1411.2335, 2014.