Reconstruction of 3D ball/shuttle position by two image points from a single view

Lejun Shen, Qing Liu, Lin Li, Yawei Ren

Chengdu Sport University, ChengDu, China,
lejun.shen@cdsu.edu.cn,

Abstract. Monocular 3D reconstruction is an important problem. We solve the problem of reconstructing a 3D ball or shuttle position from a single-view television video. The contextual constraint is vital in this paper, which is implemented by a confirming point. The confirming point represents all the contextual cues, such as human pose, stroke technique, and shadow on the ground. The confirming point also tells us where the 3D point is. The confirming point decision is made by a human operator. Thus, the proposed method is a mixture of computer vision and human intelligence. Moreover, we propose a new air-ball friction model. This model provides a more accurate result because the aerodynamic drag force cannot be ignored in ball game.

Keywords: computer vision, monocular, 3D reconstruction, badminton

1 Introduction

Monocular 3D reconstruction is an important problem and has received much interest both in the computer vision community and in the "computer science in sport" community. We consider the problem of estimating 3D ball or shuttle position from a single-view television video. Compared with multi-view methods, there are many advantages to use single-view (or monocular) method, including lower equipment cost, lower complexity, and higher portability of camera set-up. Moreover, most of the publicly available sport videos are single view, especially the data of competitors. Finally, the television sport video provides more realistic and more important information than the data from laboratory.

Monocular 3D reconstruction is a challenging area with many unsolved problems. In 2015, Shen et al. solved the model-drifting problem using a mixture of physical model and geometric model [8]. The hidden assumption of this solution is that the ball must hit the ground plane and the 3D position of ball bounce is known. However, this assumption is not valid in volleyball game or badminton game because the ball/shuttle is not allowed to bounce. We solve this problem using a mixture of computer vision and human intelligence.

Monocular 3D reconstruction is an ill-posed problem and is inherently ambiguous: a single RGB image on its own does not provide depth cue explicitly (i.e., given a single view image, there are infinite numbers of 3D scene structures explaining this image). However only one explanation is, in fact, correct. In this

paper, we present a novel method for the reconstruction of 3D ball/shuttle position in sports television using a new constraint, namely, confirming point.

2 Method

2.1 The 3D Line Constraint

A camera is a mapping between the 3D world and a 2D image. Camera mapping is represented by a matrix P. According to the pinhole camera model, P is a 3×4 projective matrix defined by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}\ m_{12}\ m_{13}\ m_{14} \\ m_{21}\ m_{22}\ m_{23}\ m_{24} \\ m_{31}\ m_{32}\ m_{33}\ m_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where $m_{11}, m_{12}, ..., m_{34}$ are the elements of matrix P, (X, Y, Z) is the position in the 3D world and (u, v) is the position in the 2D image. Let $m_{34} = 1$ because the equation (1) uses homogeneous coordinates system.

Given N ($N >= 6$) court-to-image point correspondences, the projective matrix P is computed by solving the following linear equations

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1X_1 & -v_1X_1 \\ & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NZ_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_NX_N & -v_NY_N & -v_NZ_N \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ \cdot \\ \cdot \\ u_N \\ v_N \end{bmatrix} \tag{2}$$

Suppose P is known and a 2D image point (u, v) is given, a set of 3D points maps to this image point. This set constitutes a ray or line in 3D space passing through the camera center. Monocular 3D reconstruction is very difficult because one 2D image point back projects to one ray, which is defined by two planes:

$$\begin{aligned} (um_{31} - m_{11})X + (um_{32} - m_{12})Y + (um_{33} - m_{13})Z + (um_{34} - m_{14}) = 0 \\ (vm_{31} - m_{21})X + (vm_{32} - m_{22})Y + (vm_{33} - m_{23})Z + (vm_{34} - m_{24}) = 0 \end{aligned} \tag{3}$$

As shown in Fig. 1, the image point (u, v) back project to a 3D line L. It should be emphasized that the aim of this paper is to find the 3D point (X, Y, Z) on this 3D line.

To find the 3D point, it will be necessary to add constraints on the solution, which at the beginning result from a physical model. A well-known physical constraint is gravity. Ohno et al. showed that the 3D flying ball trajectory can be

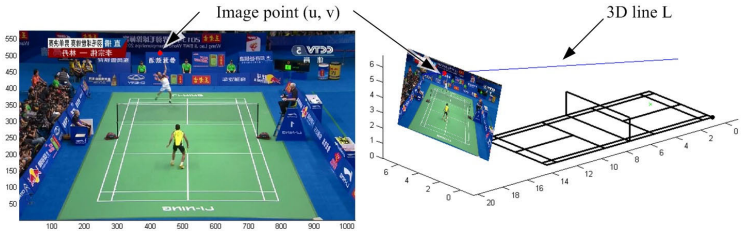

Fig. 1. Image point and its corresponding 3D line

estimated by fitting a physical model of 3D ball movements in the world to the observed 2D ball trajectory in the image [5]. Ribnick et al. proved the minimum condition for the existence of a unique solution[6]. The basic assumption behind these methods is that the ball motion is governed only by gravity. This assumption suffers from the model-drifting problem. Shen et al. added a bounce point constraint to fix this problem[8]. However, this constraint is NOT available in badminton or volleyball game because the shuttle or ball does not hit the ground plane until one player scores a point.

In this paper we will develop a new method that uses *confirming point* to reconstruct 3D shuttle position in a badminton game when the racket hits the shuttle.

2.2 Cylinder Constraint

As shown in Fig. 2, one image point back project to one ray. Suppose we have a new camera ♯2, then two matching image points back project to two rays (L and L' in Fig. 2). The 3D point (X, Y, Z) can be reconstructed by intersecting these two rays. This method is well-known as multi-view (or triangulation) method and camera ♯2 introduces a new constraint on the solution, which indicates the 3D point (X, Y, Z) on line L. Hawkeye system is a typical example.

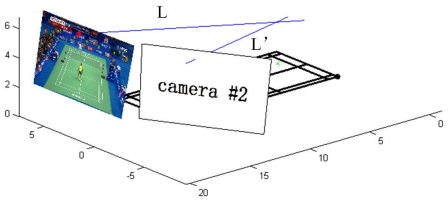

Fig. 2. Multi-view method

In monocular 3D reconstruction, however, camera ♯2 is not available. In computer vision literature, there are several other constraints, including camera motion[4], shading[9], structured lighting[7], motion blur[2], learned pattern[3], etc. However these constraints cannot be directly used in sport games. We found that contextual constraints are very useful in sport games. We give a simple example of player's positional cylinder constraint.

In a badminton game, a player hits a shuttlecock across the net using a racket. The strike-point cannot be far away from the player's body, which is modeled by a cylinder at the player's position (4.29, 1.41, 0) in Fig. 3. Humans make rough inferences about the approximate 3D point using cylinder constraint. Empirically, the radius of the cylinder is 1.84 m, and the height of the cylinder is 3.2 m. In Fig. 3, the image point (595, 68) back projects to a 3D line. After introduction of contextual constraint (player's position here), the 3D line L is reduced to a red colored line segment within the cylinder.

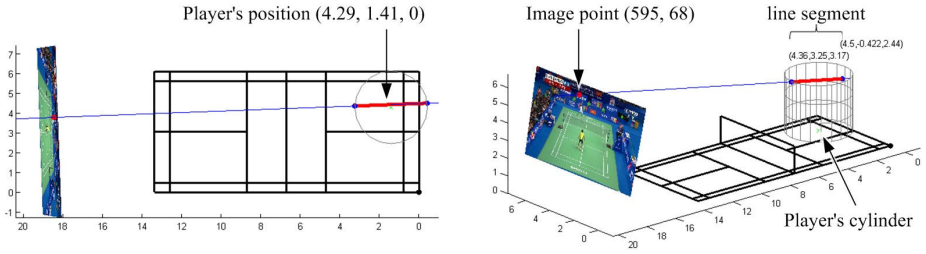

Fig. 3. Player's positional constraint

2.3 Confirming Point

The cylinder constraint has two problems: (1) it provides approximate range not accurate 3D position; (2) the player's position is ambiguous when he plays a vertical jump.

First, the more accurate 3D point can be inferred from 3D human pose when the player hits the shuttle. In other words, the contextual constraint is the human skeleton structure. Monocular 3D pose estimation has received much interest in computer vision community due to the wide range of applications. It is a difficult problem because there are large numbers of degrees of freedom (around 30) to be recovered. Especially, there are many difficulties in real sports data (e.g., motion blur, self-occlusions, and self-similarity). Therefore, fully automatic 3D pose estimation is currently not possible, and manual labeling is required [1].

The complete 3D skeleton requires a human operator to manually label all joint locations in image. These labeling is a heavy work in badminton game because the average length of a rally is more than 10 shots. Moreover, a full-body 3D human pose is not necessary if we want only the 3D position of the

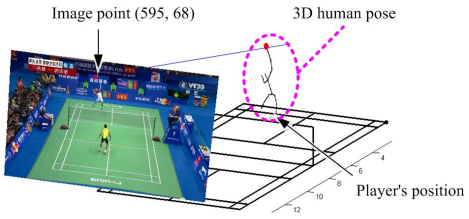

Fig. 4. The 3D point can be inferred by 3D human pose

shuttle in sports television. In this paper, we use a confirming point rather than all joint points. This confirming point can be easily labeled as follows.

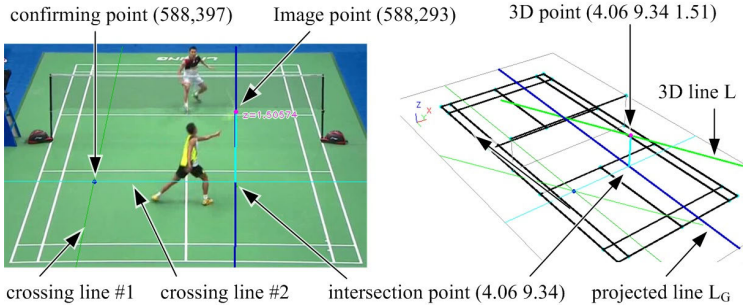

Fig. 5. The 3D point is inferred from image point and confirming point

In Fig. 5, for example, the player hits the shuttle on his backhand side in the midcourt. The operator clicks the shuttle position (588,293) in the image, and the corresponding 3D line L is computed according to (3). The operator's mouse moves on the computer screen, and the confirming point is shown both in the image (left) and in the 3D visualization window (right). The confirming point (588,397) is decided by the operator according to the contextual cue: human pose, stroke technique, shadow on the ground, and shuttle trajectory. The confirming point has two assistant crossing lines, and each line parallels to the X or Y axis. Crossing line ♯2 intersects with the projected line L_G of L on the ground plane. The intersection point (4.06 9.34) indicates that the 3D point is (4.06 9.34 1.51). In short, two 2D image points provide one 3D point.

Please note that the confirming point decision is made by human operator, because contextual cues are too complicated to be captured by mathematical model using current computer vision technology. For example, the player's position is ambiguous when he plays a vertical jump smash. The fact is that the shadow of the player indicates that he is jumping (left in Fig. 6), but the image

itself gives the incorrect position (jump height is 0 m). However, the temporal cues show that the correct position of player is not on the ground, and the vertical jump height is 0.42 m (right in Fig. 6). Therefore, a human operator can easily tell the correct confirming point, but computer can not.

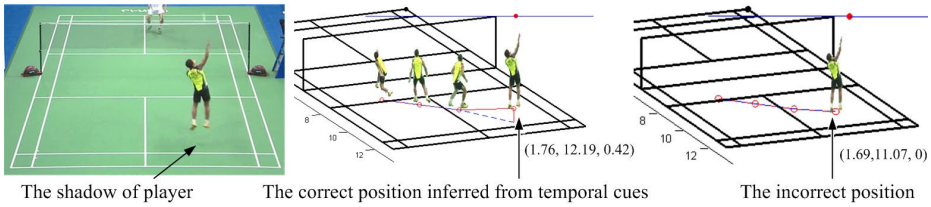

Fig. 6. Example of a vertical jump

The confirming point plays a vital role in our monocular 3D reconstruction system. It represents all of the contextual information perceived by humans, and helps us to tell the 3D position from a single view. Hence, our method is a mixture of computer vision and human intelligence.

2.4 Air-ball Friction Model

Given 3D position of ball/shuttle, complete 3D trajectory can be estimated using Shen's method [8]. But a main shortcoming of [6] [8] is that it doesn't model the air-ball friction. In this paper, we propose a novel air-ball friction model.

$$\begin{aligned} X(t) &= X(0) + tV_X - \tfrac{1}{2}\alpha V_X t^2 \\ Y(t) &= Y(0) + tV_Y - \tfrac{1}{2}\alpha V_Y t^2 \\ Z(t) &= Z(0) + tV_Z - \tfrac{1}{2}(g + \alpha V_Z)t^2 \end{aligned} \tag{4}$$

where $(X(t), Y(t), Z(t))$ denote the position of the ball at time t, and (V_X, V_Y, V_Z) denote the velocity of the ball at time 0, g denotes the acceleration of gravity ($g = 9.8\ m/s^2$) and α is *air-ball friction coefficient*. The aerodynamic drag force is a function of ball velocity (V_X, V_Y, V_Z) and air-ball friction coefficient (α). This 3D ball trajectory model is very simple and effective.

First, we can directly estimate α from a monocular view data.

$$\theta^* = \arg\min_{\theta}(F) \tag{5}$$

where $\theta = (X(0), Y(0), Z(0), V_X, V_Y, V_Z, \alpha)$ and F is a cost function[8]. But this α estimation method is troublesome. Evidence shows that the monocular view data itself cannot provide a correct result.

Moreover, we collected a Hawkeye ground truth dataset GT and $GT(i)$ is i^{th} smash speed (i=1,2,...G) in badminton game from television. We estimate α by

minimization:

$$\alpha^* = \arg\min_{\alpha} \sum_{i=1}^{G} (\sqrt{(V_X^2 + V_Y^2 + V_Z^2)} - GT(i))^2 \quad (6)$$

We found that the best air-ball friction coefficient α^* is 3.135 and the best mixture weight w is 780 in [8]. Given these two parameters, our α fixed method provides a reliable result. Please refer to section 3 for more details.

3 Experiment

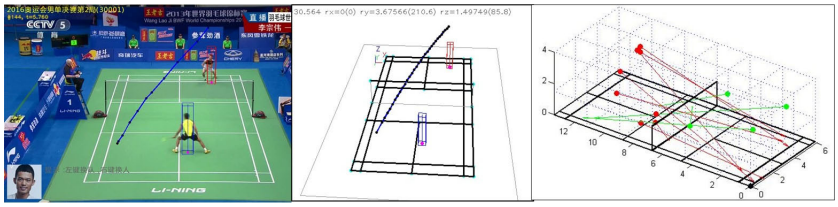

Fig. 7. Our monocular 3D reconstruction system and the final strikes

We implement the proposed method using C++ and run it on a 3.3GHz PC with 2G RAM. The videos are recorded from the Badminton - Singles Men of the Rio 2016 Olympic Games. Using the monocular 3D reconstruction method in [8] and two image points method in this paper, we obtain 1363 shuttle trajectories. Some final strikes are visualized in Fig. 7.

Table 1. The Comparison of four methods.

Method	Mean-Square Error (m/s)	Back-project Error (pixels)
Ribnick 2009 [6]	2.27e+018	**23.212**
Shen 2015 [8]	5989.14	209.828
Our α estimation method	5989.08	206.85
Our α fixed method	**1559.37**	67.0942

Second, we collected a Hawkeye ground truth dataset from the original badminton games in television broadcast, and compared four methods: Ribnick 2009 [6], Shen 2015 [8], our α estimation method and our α fixed method (α=3.135). As shown in Table 1, Ribnick 2009 is the best according to the back-project error. But it is the worst according to the MSE because it suffers from the model drifting problem [8]. Our α fixed method is the best. The experiment result shows

that the aerodynamic drag force (α) influences the monocular 3D reconstruction significantly and cannot be ignored.

4 Conclusion

The proposed method uses two constraints to solve the monocular 3D reconstruction problem. The first constraint is the ball/shuttle image point recovering a 3D line from camera to ball. The second constraint is the confirming point telling us where the 3D point is on this line. The confirming point represents all the contextual cues, such as human pose, stroke technique, and shadow on the ground. The confirming point decision is made by a human operator. Therefore, the proposed method is a mixture of computer vision and human intelligence.

The proposed method models air-ball friction to improve monocular 3D reconstruction. The aerodynamic drag force influences monocular 3D reconstruction significantly and cannot be ignored. The air-ball friction coefficient α is 3.135 in badminton game.

Acknowledgement. This work was supported by Technology Research and Development Program of Sichuan Province of China (2015JY0148).

References

1. Akhter, I., Black, M.J.: Pose-conditioned joint angle limits for 3d human pose reconstruction. In: Computer Vision and Pattern Recognition (CVPR). pp. 1446–1455 (2015)
2. Boracchi, G., Caglioti, V., Giusti, A.: Single-image 3d reconstruction of ball velocity and spin from motion blur. In: The 3rd International Conference on Computer Vision Theory and Applications. vol. 22, pp. 22–29 (2008)
3. Karsch, K., Liu, C., Kang, S.B.: Depth extraction from video using non-parametric sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(11), 775–788 (2014)
4. Lu, Y., Zhang, J.Z., Wu, Q.M.J., Li, Z.N.: A survey of motion-parallax-based 3-d reconstruction algorithms. IEEE Transactions on Systems Man and Cybernetics Part C 34(4), 532–548 (2004)
5. Ohno, Y., Miura, J., Shirai, Y.: Tracking players and estimation of the 3d position of a ball in soccer games. In: International Conference on Pattern Recognition (ICPR). vol. 1, pp. 145–148. IEEE (2000)
6. Ribnick, E., Atev, S., Papanikolopoulos, N.P.: Estimating 3d positions and velocities of projectiles from monocular views. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(5), 938–944 (2009)
7. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recognition 43(8), 2666–2680 (2010)
8. Shen, L., Liu, Q., Li, L., Yue, H.: 3d reconstruction of ball trajectory from a single camera in the ball game. In: International Symposium on Computer Science in Sports (ISCSS). pp. 33–39. Springer (2015)
9. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from shading: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(8), 690–706 (1999)