3D Ball Trajectory Reconstruction from Single-Camera Sports Video For Free Viewpoint Virtual Replay

Hua-Tsung Chen
Information & Communications Technology Lab
National Chiao Tung University
Hsinchu, Taiwan
huatsung.chen@gmail.com

Chien-Li Chou, Wen-Jiin Tsai, Suh-Yin Lee
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
{fallwind, wjtsai, sylee}@cs.nctu.edu.tw

***Abstract*—Free viewpoint video presentation is a new challenge in multimedia analysis. This paper presents an innovative physics-based scheme to reconstruct the 3D ball trajectory from single-camera volleyball video sequences for free viewpoint virtual replay. The problem of 2D-to-3D inference is arduous due to the loss of 3D information in projection to 2D images. The proposed scheme incorporates the domain knowledge of court specification and the physical characteristics of ball motion to accomplish the 2D-to-3D inference. Motion equations with the parameters are set up to define the 3D trajectories based on physical characteristics. Utilizing the geometric transformation of camera calibration, the 2D ball coordinates extracted over frames are used to approximate the parameters of the 3D motion equations, and finally the 3D ball trajectory can be reconstructed from single-camera sequences. The experiments show promising results. The reconstructed 3D trajectory enables the free viewpoint virtual replay and enriched visual presentation, making game watching a whole new experience.**

I. INTRODUCTION

The proliferation of multimedia data necessitates the development of automatic systems and tools for content-based multimedia information retrieval. Due to the commercial benefits and audience requirement, sports video analysis has been attracting considerable attention. Various approaches of shot classification [1, 2], highlight extraction [3, 4] and semantic annotation [5, 6, 7] are developed by fusing audiovisual features with game-specific rules.

Some work focuses on 3D trajectory reconstruction based on multiple cameras located at specific positions [8-10]. Hawk-Eye system produces computer-generated replays viewed through 360 degrees [8]. 2D tracking is first performed on each of the specifically located cameras. These 2D trajectories are then sent to a 3D reconstitution module to construct the 3D trajectories, and impact points between separate trajectories (can occur at a bounce or a strike) are determined. Finally, the complete track is visualized. ESPN K-Zone system extracts the trajectory for each pitch and uses computer-generated graphics to outline the strike zone boundaries [9]. Two cameras linked to two PCs are used to observe the ball and each PC extracts a 2D trajectory. The two pitch-tracking computers combine two 2D positions which correspond to the same time code into a 3D position. Then, the successive 3D positions are fed into a Kalman filter to determine the final trajectory. Umpire Information System uses multiple cameras to track each pitch and measure the batter's strike zone for supporting the strike/ball judgment [10]. However, the high demand for the installation locations and the visible areas of the multiple cameras constrains the practicability of their systems.

More keenly than ever, the audience desires the freedom of choosing the viewpoint at will [11, 12]. Traditional interactive video viewing systems for quick browsing, indexing and summarization no longer fulfill their requirements. Hence, in this paper we propose a physics-based scheme for 3D trajectory reconstruction from single-camera volleyball sequences to provide the free viewpoint virtual replay of the ball motion, present informative insights into the game and make game watching an entirely novel experience. Fig. 1 illustrates the system framework.

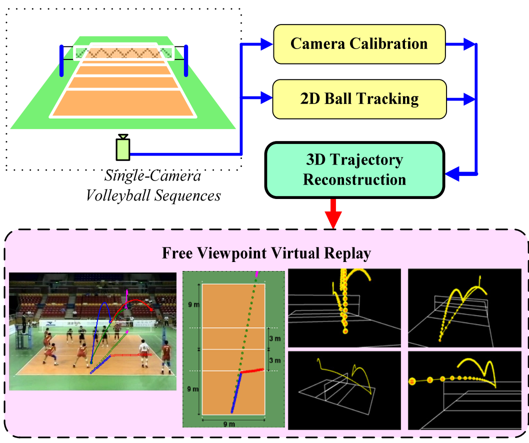


Figure 1. Overview of the system framework

This work is particularly supported by "Aim for the Top University Plan" of the National Chiao Tung University and Ministry of Education, Taiwan, R.O.C. and is partially supported by National Science Council of R.O.C. under the grant no. 98-2221-E-009-091-MY3.

We first perform camera calibration via finding the non-coplanar feature points to compute the projection matrix mapping 3D real world coordinates to 2D image positions. For 2D ball tacking, our previous work [13] is applied. Ball candidates are detected in each frame under the constraints of size, shape and compactness. Then, we correlate motion information on the ball candidates over a sequence of frames, explore potential trajectories and identify the true ball trajectories.

For 3D trajectory reconstruction, we set up the motion equations with the parameters (velocities and initial positions) to model the 3D trajectory based on the physical characteristics of ball motion. The 3D ball positions over frames are represented by equations. The projection matrix computed in camera calibration is then used to map the equation-represented 3D ball positions to the 2D ball coordinates in frames. With the 2D coordinates of the extracted ball candidates being known, we can compute the parameters of the 3D motion equations. Finally, the 3D positions and velocities of the ball can be derived to reconstruct the 3D ball trajectory. Needless of multiple cameras specifically located, the proposed system can present free viewpoint virtual replay, enabling the audience to watch the ball motion from arbitrary perspectives.

The rest of this paper is organized as follows. Sec. II describes the process of camera calibration. Sec. III presents the proposed 3D trajectory reconstruction and Sec. IV shows the simulation results. Finally, Sec. V concludes this paper.

II. Camera Calibration

Camera calibration provides geometric transformation mapping between the 3D real world and the 2D image space. We use W for the real world point represented by a homogeneous 4-vector $(X, Y, Z, 1)^{\mathrm{T}}$, m for the image point represented by a homogeneous 3-vector $(x, y, 1)^{\mathrm{T}}$, and P for the 3×4 homogeneous camera projection matrix. Then, the mapping between the 3D real world and the 2D image is written compactly as Eq. (1).

$$m = \mathrm{P}W \tag{1}$$

To compute the matrix P, we need to extract a set of *corresponding points*—the points whose coordinates are both known in the 3D real world and in the 2D image. The court region consisting of the court lines L_1 to L_7 (see Fig. 2) is first segmented by the dominant color feature. Then, we detect the court lines by the Hough transform and obtain the coordinates of the ground feature points g_1 to g_{10} via computing the intersection of court lines. In addition to the ground feature points, the computation of camera matrix requires non-coplanar feature points. Thus, we trace vertically from the ground points g_5, g_6 in the image and search the two vertical borders of the net using the Hough transform. The endpoints of the vertical border of the net (p_1 to p_4), together with the ground feature points, form a non-coplanar feature point set.

Here, we briefly describe the methods of solving for the camera projection matrix P; the details of these methods are available in [14]. For each correspondence W ↔ m we derive a relationship from Eq. (1) as Eq. (2):

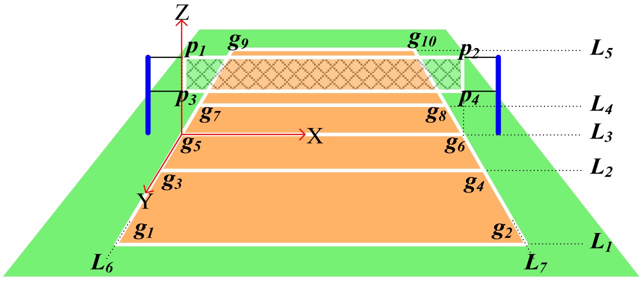


Figure 2. Illustration of the non-coplanar feature points.

$$\begin{bmatrix} 0^T & -W^T & yW^T \\ W^T & 0^T & -xW^T \\ -yW^T & xW^T & 0^T \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \tag{2}$$

where each $P^{i\mathrm{T}}$ is a 4-vector, the i-th row of P. Although there are three equations in (2), only two of them are linear independent (since the third row is obtained, up to scale, from the sum of x times the first row and y times the second). Thus each point correspondence gives two equations in the entries of P. It is usual to choose only the first two equations:

$$\begin{bmatrix} 0^T & -W^T & yW^T \\ W^T & 0^T & -xW^T \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \tag{3}$$

Since the matrix P has 12 entries, and (ignoring scale) 11 degrees of freedom, it is necessary to have at least 11 equations to solve for P. With the 14 non-coplanar point correspondences obtained, we can solve for P using the direct linear transform (DLT) algorithm (see p109, p178-184 in [14]).

III. 3D Trajectory Reconstruction

In volleyball games, the ball trajectory comprises a sequence of near parabolic curves, even though many factors affect the ball motion, such as velocity, gravity acceleration, spin axis, spin rate, air friction, etc. We call each near parabolic curve in the ball trajectory a *sub-trajectory* and roughly model a 3D sub-trajectory as:

$$\begin{aligned} X_t &= X_0 + V_X\, t \\ Y_t &= Y_0 + V_Y\, t \\ Z_t &= Z_0 + V_Z\, t + gt^2/2 \end{aligned} \tag{4}$$

where (X_t, Y_t, Z_t) is the 3D ball coordinate at time t, (X_0, Y_0, Z_0) is the 3D ball coordinate of the starting position in the sub-trajectory, (V_X, V_Y, V_Z) is the 3D ball velocity and g is the gravity acceleration. We use $W_t = (X_t, Y_t, Z_t, 1)^{\mathrm{T}} = (X_0 + V_X\, t, Y_0 + V_Y\, t, Z_0 + V_Z\, t + gt^2/2)^{\mathrm{T}}$ and $m_t = (x_t, y_t, 1)^{\mathrm{T}}$. From Eq. (3) each point correspondence gives two equations as

$$\begin{bmatrix} 0^T & -W_t^T & y_t W_t^T \\ W_t^T & 0^T & -x_t W_t^T \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \tag{5}$$

Given N detected ball candidates, we obtain $2N$ equations. Since the entries in P, the 2D image coordinate (x_t, y_t) and the occurring time t of each ball candidate are known, we can set up a linear system $M^{A}_{2N\mathrm{x}6}\, M^{B}_{6\mathrm{x}1} = M^{C}_{2N\mathrm{x}1}$, where $M^{B} = (X_0, Y_0, Z_0, V_X, V_Y, V_Z)^{T}$, to compute the six unknowns X_0, Y_0, Z_0, V_X, V_Y and V_Z. We solve for M^B using the direct linear transform algorithm (see p109 in [14]). Thus, each 3D sub-trajectory can be reconstructed.

However, the experiments shows that the 3D coordinate of the turning point between two adjacent sub-trajectories computed from the preceding sub-trajectory is not always consistent with the one computed from the succeeding sub-trajectory. To overcome this problem, we enhance the algorithm by taking two adjacent sub-trajectories into consideration simultaneously.

Fig. 3 illustrates the procedure of 3D trajectory reconstruction by a sample trajectory. The ball trajectory contains three sub-trajectories S_0, S_1 and S_2, as shown in Fig. 3(a). Let P_1 be the turning point between S_0 and S_1, and P_2 between S_1 and S_2. Let (V_{Xi}, V_{Yi}, V_{Zi}) be the 3D ball velocity of the sub-trajectory S_i, where i is the index of the sub-trajectories. As shown in Fig. 3(b), we consider the two adjacent sub-trajectories S_0 and S_1 to derive $(V_{X0}, V_{Y0}, V_{Z0}, X_1, Y_1, Z_1, V_{X1}, V_{Y1}, V_{Z1})$. Taking P_1 as the starting point, the 3D sub-trajectories of S_0 and S_1 are expressed as Eq. (6) and Eq. (7), respectively:

$$\begin{aligned} X_t &= X_1 - V_{X0}\, t \\ Y_t &= Y_1 - V_{Y0}\, t \\ Z_t &= Z_1 - V_{Z0}\, t + gt^2/2 \end{aligned} \qquad (6)$$

$$\begin{aligned} X_t &= X_1 + V_{X1}\, t \\ Y_t &= Y_1 + V_{Y1}\, t \\ Z_t &= Z_1 + V_{Z1}\, t + gt^2/2 \end{aligned} \qquad (7)$$

Using $W_t = (X_t, Y_t, Z_t, 1)^T = (X_1 - V_{X0}\, t, Y_1 - V_{Y0}\, t, Z_1 - V_{Z0}\, t - g\, t^2/2)$ for S_0 and $W_t = (X_t, Y_t, Z_t, 1)^T = (X_1 + V_{X1}\, t, Y_1 + V_{Y1}\, t, Z_1 + V_{Z1}\, t + g\, t^2/2)$ for S_1, we obtain two equations for each ball candidate, as Eq. (5). The $2N$ equations produced from N ball candidates form a linear system: $M^{D}_{2N\mathrm{x}9}\, M^{E}_{9\mathrm{x}1} = M^{F}_{2N\mathrm{x}1}$, where $M^{E} = (V_{x0}, V_{y0}, V_{z0}, X_1, Y_1, Z_1, V_{x1}, V_{y1}, V_{z1})^T$. We solve for M^E using the direct linear transform algorithm. Thus, the coordinate of P_1 (X_1, Y_1, Z_1) is obtained. In the same way, the nine parameters $(V_{X1}, V_{Y1}, V_{Z1}, X_2, Y_2, Z_2, V_{X2}, V_{Y2}, V_{Z2})$ can be computed by processing S_1 and S_2 simultaneously, as shown in Fig. 3(c).

For the sub-trajectory S_1 between two turning points P_1 and P_2, its 3D velocity (V_{X1}, V_{Y1}, V_{Z1}) is computed twice: one when processing S_0-P_1-S_1 and the other when processing S_1-P_2-S_2. Thus, we refine (V_{X1}, V_{Y1}, V_{Z1}) via solving for Eq. (7) with $(X_t, Y_t, Z_t) = (X_2, Y_2, Z_2)$. Finally, the 3D trajectory can be reconstructed with the 3D ball velocity (V_{Xi}, V_{Yi}, V_{Zi}) on each sub-trajectory S_i and the coordinate (X_j, Y_j, Z_j) of each turning point P_j.

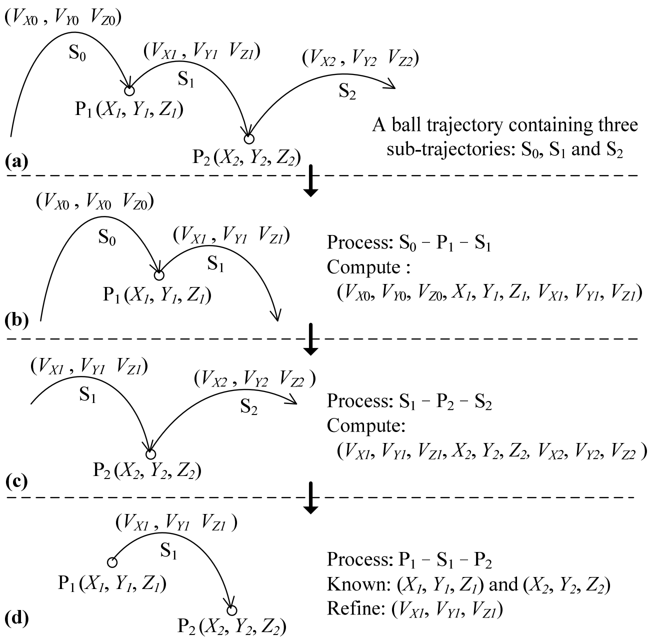
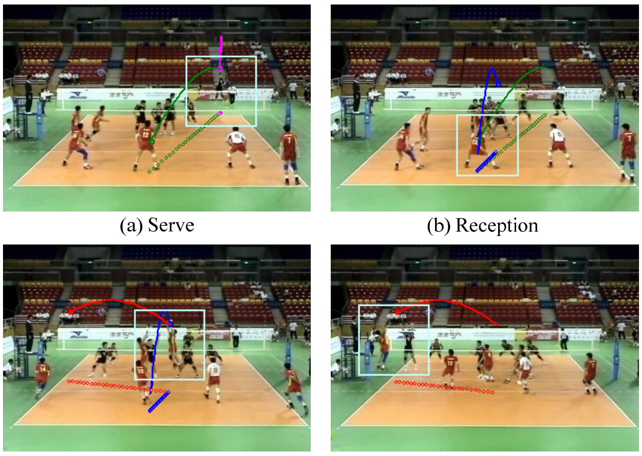


Figure 3. Procedure of 3D trajectory reconstruction.

IV. SIMULATION RESULTS

The proposed scheme has been tested on the 92 volleyball video sequences (MPEG-1, 352×240, 29.97 fps) captured in the 2005 Asia Men's Volleyball Challenge Cup. Since the ground truth of 3D ball positions can hardly be obtained, the results of 3D trajectory reconstruction are evaluated subjectively by manual inspection. Among the 92 sequences, we achieve good reconstruction in 77 sequences (the accuracy is about 84%). Due to the limited space, we present the results for only one sequence. Fig. 4 displays the enriched frames at the moments of serve, reception, set and spike. Each frame is enriched by superimposing the ball trajectory on the frame and projecting the 3D trajectory on the court plane. It can be observed that the turning points of the ball trajectory are just the positions of the actions occurring. Fig. 5(a) presents the free viewpoint virtual replay from various perspectives, which can be switched flexibly. In Fig. 5(b), the trajectory projected on the court model visualizes the transition of ball motion.

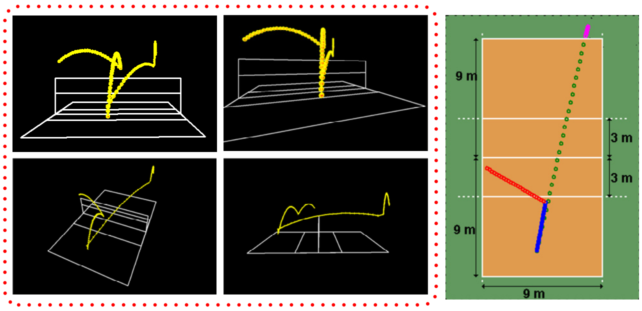
Inspecting the error cases, we find that the error in 2D ball tracking might cause the system to misjudge a far-to-near trajectory as a near-to-far one, and vice versa, as shown in Fig. 6. In this case, the 3D ball positions could not be computed correctly. Strictly speaking, there may be some deviation between the actual ball trajectory and the reconstructed 3D trajectory, due to the effects of the physical factors we do not involve, such as air friction, ball spin rate and spin axis, etc. However, our experimental results show that the proposed physics-based method is able to reconstruct the 3D ball trajectory pretty well for providing the free viewpoint virtual replay and the enriched visual presentation.



(a) Serve (b) Reception

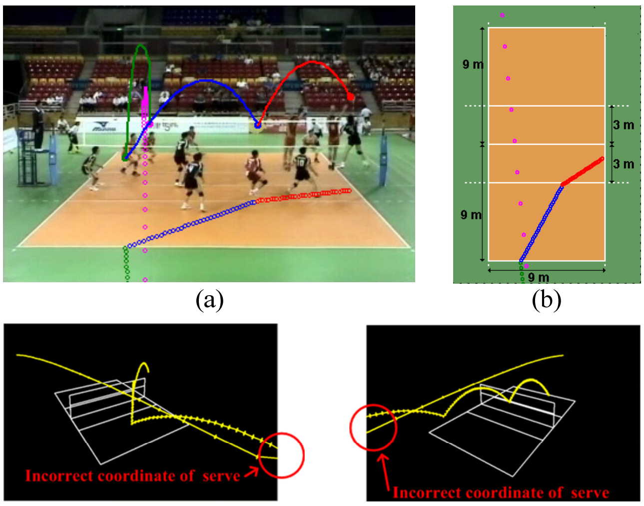
(c) Set (d) Spike

Figure 4. Enriched frames for serve, reception, set and spike.



(a) (b)

Figure 5. Free viewpoint virtual replay. (b) Visualization of ball motion projected on the court model.



(a) (b)

(c) (d)

Figure 6. Example of false 3D trajectory reconstruction: (a) Enriched frame, (b) Visualization of ball motion projected on the court model, (c) and (d) 3D virtual replays from different viewpoints.

V. CONCLUSIONS

In this paper, we propose an innovative and practical scheme to reconstruct the 3D ball trajectory from single-camera sequences by utilizing the domain knowledge of court specification for camera calibration and the physical characteristics of ball motion for 3D information inference. Based on the reconstructed 3D trajectory, the free viewpoint virtual replay and the enriched visual presentation can be presented to the audience needless of multiple specifically located cameras. An entirely new and exciting experience of game watching can be achieved.

Based on the proposed 3D trajectory reconstruction scheme, the possible applications are manifold, such as serve placement prediction, ball speed estimation, set curve analysis, etc. In the future, we want to develop an automatic game study system for the visualization of the semantic content and providing insights into the performance, style and strategy of the game.

REFERENCES

[1] L.-Y. Duan, M. Xu, and Q. Tian, “A unified framework for semantic shot classification in sports video,” *IEEE Trans. on Multimedia*, vol. 7, pp. 1066-1083, Dec. 2005.

[2] M.-C. Tien, H.-T. Chen, Y.-W. Chen, M.-H. Hsiao and S.-Y. Lee, “Shot classification of basketball videos and its applications in shooting position extraction,” in *Proc. IEEE ICASSP*, pp. I-1085-1088, 2007.

[3] C.-C. Cheng, and C.-T. Hsu, “Fusion of audio and motion information on HMM-based highlight extraction for baseball games,” *IEEE Trans. on Multimedia*, vol.8, pp. 585-599, Jun. 2006.

[4] H.-S. Chen, H.-T. Chen, W.-J. Tsai, S.-Y.Lee and J.-Y. Yu, “Pitch-By-Pitch Extraction from Single View Baseball Video Sequences,” in *Proc. ICME*, pp. 1423-1426, 2007.

[5] J. Assfalg, M. Bertini, C. Colombo, A. Del Bimbo and W. Nunziati, “Semantic annotation of soccer videos: automatic highlights identification,” *Computer Vision and Image Understanding*, vol. 92 (2-3), pp. 285-305, 2003.

[6] H.-T. Chen, C.-L. Chou, W.-J. Tsai, S.-Y. Lee and J.-Y. Yu “Extraction and Representation of Human Body for Pitching Style Recognition in Broadcast Baseball Video,” in *Proc. IEEE ICME*, 2011.

[7] H.-T. Chen, M.-H. Hsiao, H.-S. Chen, W.-J. Tsai and S.-Y. Lee, “A baseball exploration system using spatial pattern recognition,” in *Proc. IEEE ISCAS*, pp. 3522-3525, 2008

[8] N. Owens, C. Harris, and C. Stennett, “Hawk-eye tennis system,” in *Proc. VIE 2003*, pp.182- 185, 2003.

[9] A. Gueziec, “Tracking pitches for broadcast television,” *Computer*, vol.35, pp. 38-43, Mar. 2002.

[10] QuesTec – Umpire Information System. [Online].Available: http://www.questec.com/q2001/prod_uis.htm

[11] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, “3D video and free viewpoint video - technologies, applications and MPEG standards,” in *Proc. IEEE ICME*, pp.2161-2164, 2006.

[12] J. Carranza, C. Theobalt, MA Magnor, H.-P. Seidel, “Free-viewpoint video of human actors,” *ACM Trans. on Graphics*, vol. 22(3), pp. 569-577, 2003.

[13] H.-T. Chen, H.-S. Chen and S.-Y. Lee, “Physics-based ball tracking in volleyball videos with its applications to set type recognition and action detection,” in *Proc. IEEE ICASSP*, pp. I- 1097-1100, 2007.

[14] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press 2003(2nd ed.), UK.