ESTIMATING THE SPIN OF A TABLE TENNIS BALL USING INVERSE COMPOSITIONAL IMAGE ALIGNMENT

Toru Tamaki⋆, *Haoming Wang*⋆, *Bisser Raytchev*⋆, *Kazufumi Kaneda*⋆, *Yukihiko Ushiyama*†

⋆ Hiroshima University † Niigata University

ABSTRACT

In this paper, we propose a method for estimating the rotational velocity of a table tennis ball with Inverse Compositional Image Alignment, ICIA. Assuming orthogonal projection of the camera, we derive an update rule for the motion parameters. Because of the precomputation of the Hessian matrix in ICIA and the simplifying assumptions, the motion parameters estimation at each frame is very fast. We evaluate results for measuring the spins in a synthesized sequence, as well as in real image sequences of table tennis rallies.

Index Terms— table tennis, Inverse Compositional Image Alignment, ball spin estimation, rotation measurement

1. INTRODUCTION

In this paper, we propose a method for measuring the spin of a table tennis ball. In table tennis, the spin of the ball is one of the important indications used to evaluate the skills of a player. For a skillful player, the spin exceeds 5000 rotations per minute (rpm), while it is typically around 3000 rpm for a novice [3, 4]. Presently, the spin is measured either by using a spinometer or by performing a 2D image analysis with a high-speed camera. For such an application, the use of computer vision techniques can be very useful.

Tracking the motion of a rigid object, such as a ball, is a problem which has been studied since the 1980s. Early works were based on calculating optical flow [2], while recent approaches seem to favor the use of local features [1]. However, the small size of a table tennis ball, and motion blur (see Fig. 1), make it difficult to compute optical flow, or to detect reliably features in the ball region. Assuming that clear images can be captured, several methods for ball spin measurement of sport sequences have been proposed, such as using markers [8, 9], synthesized appearance of a 3D CG model [10], motion blur [5, 6], edge-based ICP [11], physical model of ball motion [12]. For general conditions, image registration-based methods have been proposed, however, inefficient optimization [13] or exhaustive search [7] are used.

Here we propose a novel image registration-based method for measuring the spin of a ball. The main contribution of this paper lies in developing a spin measurement method based on Inverse Compositional Image Alignment (ICIA) [14, 15] that

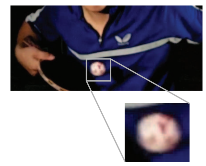


Fig. 1. An image of a table tennis player and a ball (with enlarged image).

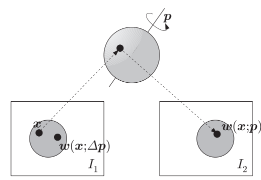


Fig. 2. Two frames relation.

accelerates computation by precomputing the Hessian matrix. Moreover, we employ several assumptions to simplify the problem formulation: the camera is orthographic and ball's radius and location at each frame are given. Also, the 3D shape of the ball is used to obtain depth information. This is similar to 2.5D ICIA [16] or 3D face alignment [17], but specialized to a sphere.

The organization of the paper is as follows. In section 2, we briefly review ICIA. Then, we introduce several assumptions necessary to simplify the problem formulation, define the motion parameters, and derive an update rule for their estimation. In section 4, we show some experimental results obtained by our prototype system for measuring spins in real image sequences of table tennis rallies.

2. INVERSE COMPOSITIONAL IMAGE ALIGNMENT

In this section, we describe ICIA in brief. We have two successive images I_1 and I_2, and let $I_1(\boldsymbol{x})$ be the intensity value at location $\boldsymbol{x} = (x, y)^T$ in image I_1. A registration minimizes the sum of squared difference between corresponding intensities in I_1 and I_2.

In ICIA, the objective function f to be minimized is given in the following form:

$$f = \sum_{\boldsymbol{x}} \left| I_1(\boldsymbol{w}(\boldsymbol{x}; \Delta \boldsymbol{p})) - I_2(\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p})) \right|^2, \tag{1}$$

where $\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p})$ is a warping function that gives the location where a point $\boldsymbol{x}$ is moved by a motion parameter $\boldsymbol{p}$. Note that $\boldsymbol{w}$ should be an identity mapping when the parameter is zero: $\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{0}) = \boldsymbol{x}$. $\Delta \boldsymbol{p}$ is obtained at each iteration step, then an update rule for ICIA $\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p}) \circ \boldsymbol{w}(\boldsymbol{x}; \Delta \boldsymbol{p})^{-1} \rightarrow \boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p})$ is used. Here $\circ$ denotes a composition of two warps that means that the warping functions should form a group: if $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are valid warp functions, so is $\boldsymbol{w}_1 \circ \boldsymbol{w}_2$.

$\Delta \boldsymbol{p}$ is calculated at each step of iterations. By using the first order Taylor expansion of the term I_1, f can be approximated as follows:

$$f = \sum_{\boldsymbol{x}} \left| I_1(\boldsymbol{x}) + \nabla I_1(\boldsymbol{x}) \frac{\partial \boldsymbol{w}(\boldsymbol{x}; \boldsymbol{0})}{\partial \boldsymbol{p}} \Delta \boldsymbol{p} - I_2(\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p})) \right|^2. \tag{2}$$

To obtain the best update $\Delta \boldsymbol{p}$ of the parameter that gives the minimum of the function, we solve $\frac{\partial f}{\partial \Delta \boldsymbol{p}} = \boldsymbol{0}$. Then, we have

$$\Delta \boldsymbol{p} = H^{-1} \sum_{\boldsymbol{x}} \left[\nabla I_1(\boldsymbol{x}) \frac{\partial \boldsymbol{w}(\boldsymbol{x}; \boldsymbol{0})}{\partial \boldsymbol{p}} \right]^T \left[I_2(\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p})) - I_1(\boldsymbol{x}) \right], \tag{3}$$

where the Hessian H is given by

$$H = \sum_{\boldsymbol{x}} \left[\nabla I_1(\boldsymbol{x}) \frac{\partial \boldsymbol{w}(\boldsymbol{x}; \boldsymbol{0})}{\partial \boldsymbol{p}} \right]^T \left[\nabla I_1(\boldsymbol{x}) \frac{\partial \boldsymbol{w}(\boldsymbol{x}; \boldsymbol{0})}{\partial \boldsymbol{p}} \right]. \tag{4}$$

The advantage of ICIA is to avoid the computation of H in each iteration step. Because computing H is very time consuming, precomputation and reuse of H make the algorithm very fast.

3. DESIGN OF THE SYSTEM

In this section, we describe the assumptions, motion parameters, and the warping function used in our system for measuring the table tennis ball's spin.

3.1. Assumptions

We assume that the camera is orthographic. This is a reasonable assumption because the ball is much smaller (40 mm in diameter) than the distance to the camera. In the experiments, the images are taken at a distance of 3 to 5 meters from a player. The ratio of the ball diameter and the distance is 1.3% to 0.8%, and then the appearance change of the ball in the image screen is less than 1% (we omit the derivation of this simple geometry due to the space limit). Therefore, the depth of the ball can be ignored.

The radius r of the ball is assumed to be given. In the experimental setup, the direction of the motion of the ball between the players is usually perpendicular to the optical axis of the camera. Hence, the change of appearance of the ball is negligible (also this is obvious from the discussion above), and it is reasonable to assume that r is constant for all frames and given in advance.

Also we assume that the center location $\boldsymbol{c} = (c_x, c_y)^T$ of the ball in every frame is given by some other source: e.g. by using circle detection with the Hough transform, or simply by user interaction.

3.2. Motion parameters

We choose the angular velocities about three axes to parameterize the spin of the ball. Angular velocities are represented by the Euler angles α, β, γ between successive frames. Then, a rotation matrix R is constructed by the angles: $R = R_\alpha R_\beta R_\gamma$. However, note that the angles are not accumulated over frames, but reset every frame to represent angular velocities (not angles) in order to avoid error accumulation.

As we consider the translation of the ball only in the image plane, translation is represented by a vector $\boldsymbol{t} = (t_x, t_y)^T$. Because the direction of the ball's motion is approximately perpendicular to the optical axis and the depth of the ball is relatively small as mentioned above, we can ignore the translation along the depth direction.

Therefore, the motion parameter vector $\boldsymbol{p}$ we use includes the following five parameters: $\boldsymbol{p} = (\alpha, \beta, \gamma, t_x, t_y)^T$.

3.3. Ball shape and depth

Since the motion parameters represent three-dimensional rigid motion, we also use the depth of the ball to register the images. The depth is obtained as follows. We assume that the target is a sphere with radius r: $x^2 + y^2 + z^2 = r^2$. Rewriting this equation, we obtain an equation for the depth at location x on the ball:

$$z(\boldsymbol{x}) = \pm \sqrt{r^2 - (x^2 + y^2)}. \tag{5}$$

We simply use the positive values of the above equation. This enables a visibility test easy because if the depth of a transformed point becomes negative, it is not visible to the camera.

3.4. The warping function

We define the warping function as follows:

$$\boldsymbol{w}(\boldsymbol{x}; \boldsymbol{p}) = P_o \left[R \begin{pmatrix} \boldsymbol{x} - \boldsymbol{c} \\ z(\boldsymbol{x} - \boldsymbol{c}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{t} \\ 0 \end{pmatrix} \right] + \boldsymbol{c}, \tag{6}$$

where z is obtained from Eq. (5), and P_o is an orthographic projection matrix:

$$P_o = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{7}$$

The warp includes a 3D rigid transformation (rotation and translation) of a 2D point $\boldsymbol{x}$. First, $\boldsymbol{x}$ is centered at $\boldsymbol{c}$ (i.e., $\boldsymbol{x}-\boldsymbol{c}$), and the z component is added to make a 3D vector. Next, it is rotated by R and translated by $\boldsymbol{t}$. Then, the 3D vector is projected to 2D by P_o. Finally, $\boldsymbol{c}$ is added to move it back.

3.5. Update rule for the motion parameters

An update rule in the form $\boldsymbol{w}(\boldsymbol{x};\boldsymbol{p}) \circ \boldsymbol{w}(\boldsymbol{x};\Delta\boldsymbol{p})^{-1} \rightarrow \boldsymbol{w}(\boldsymbol{x};\boldsymbol{p})$ is derived as follows.

Let two 2D points and their corresponding 3D points be:

$$\boldsymbol{x}_1 = \boldsymbol{w}(\boldsymbol{x};\Delta\boldsymbol{p}) = P_o\boldsymbol{X}_1, \tag{8}$$

$$\boldsymbol{x}_2 = \boldsymbol{w}(\boldsymbol{x};\boldsymbol{p}) \quad = P_o\boldsymbol{X}_2. \tag{9}$$

Now the composition of two warps is obtained by writing $\boldsymbol{x}_2$ in terms of $\boldsymbol{x}_1$ by traversing $\boldsymbol{x}_1$, $\boldsymbol{x}$, and then $\boldsymbol{x}_2$.

Here $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are as follows:

$$\boldsymbol{X}_1 = \Delta R(\boldsymbol{X}-\boldsymbol{C}) + \Delta\boldsymbol{T} + \boldsymbol{C}, \tag{10}$$

$$\boldsymbol{X}_2 = R(\boldsymbol{X}-\boldsymbol{C}) + \boldsymbol{T} + \boldsymbol{C}, \tag{11}$$

where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x} \\ z(\boldsymbol{x}-\boldsymbol{c}) \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{0} \end{pmatrix}, \boldsymbol{T} = \begin{pmatrix} \boldsymbol{t} \\ \boldsymbol{0} \end{pmatrix}. \tag{12}$$

ΔR and $\Delta\boldsymbol{T}$ are parameters corresponding to $\Delta\boldsymbol{p}$.

First, we have

$$\boldsymbol{X} = \Delta R^{-1}(\boldsymbol{X}_1 - \boldsymbol{C} - \Delta\boldsymbol{T}) + \boldsymbol{C}, \tag{13}$$

then after substituting $\boldsymbol{X}$ in $\boldsymbol{X}_2$:

$$\boldsymbol{X}_2 = (R\Delta R^{-1})(\boldsymbol{X}_1 - \boldsymbol{C}) + (\boldsymbol{T} - R\Delta R^{-1}\Delta\boldsymbol{T}) + \boldsymbol{C}. \tag{14}$$

Thus, we obtain the updated motion parameters:

$$R \leftarrow R\Delta R^{-1}, \qquad \boldsymbol{t} \leftarrow P_o(\boldsymbol{T} - R\Delta R^{-1}\Delta\boldsymbol{T}). \tag{15}$$

4. EXPERIMENTAL RESULTS

4.1. Simulation

First we evaluated the accuracy of the estimated rotation angles for a synthesized image sequence with a texture-mapped 3D sphere model. At each frame, the ball rotates by ± 5 degrees about either Y axis (first 20 frames), X axis (next 20 frames) or Z axis (last 20 frames). Fig. 3 shows the estimated rotation angles. The total spin angle was computed by $\cos^{-1}\left(\frac{\mathrm{tr}(\hat{R})-1}{2}\right)$, where $\hat{R}$ is the rotation matrix obtained from the estimated angles. RMSE of the total spin angle over 60 frames of the sequence is 0.0177 [deg] which is quite small and shows the accuracy of the proposed method.

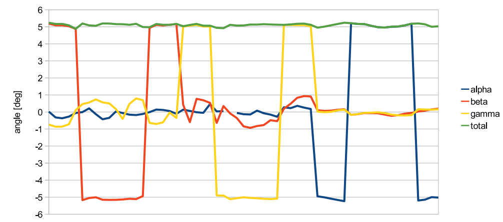
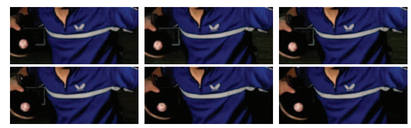
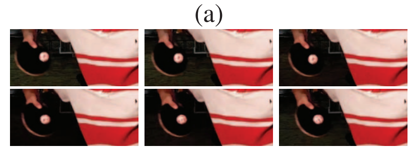


Fig. 3. Estimated rotation angles for the synthesized sequence.



(a)



(b)

Fig. 4. Sequences used in the experiments. (a) and (b) show two sequences for different players.

4.2. Real sequences

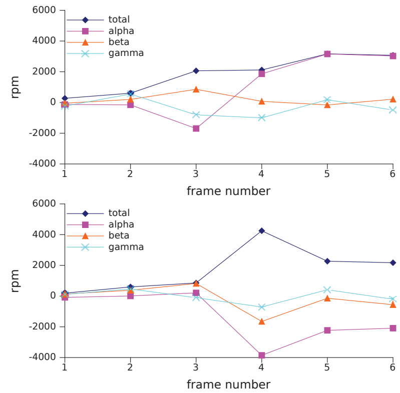
Here we describe experimental results of spin estimation for real image sequences of table tennis rallies.

Fig. 4 shows two image sequences of different players. Images were taken at 600 fps by a fixed high-speed camera with halogen lamps mounted from the side of the player. An official 40 mm table tennis ball was randomly textured by marker pens. The angle parameters α, β, and γ were measured in radian per 1/600 second because they were estimated by using two successive frames. Then the angles were converted to rotation-per-minute (rpm). The radius and center locations were obtained by manually in these experiments.

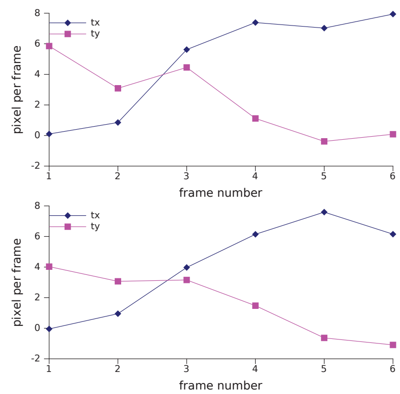
Fig. 5(a) shows the estimated spins in rpm. The spins α, β, and γ of are shown separately, while the total spin is also shown. Both sequences start one or two frames before the racket hits the ball. As can be seen, the spins are small in the first two frames, then increase rapidly at the third frame.

2D translations are shown in Fig. 5(b). The vertical axis represents translations at each frame; e.g., velocities of the ball. At the beginning of the sequences, the balls are falling vertically, then hit by the racket. Therefore, t_y is large at first (because y axis is downward), then t_x increases (x axis is rightward) at the third frame.

The accuracy of the estimates has not been evaluated yet, however, the results are very useful for players and coaches who use the system for their training because they can see the ball spins of the players quantitatively instead of impact



(a)



(b)

Fig. 5. Estimated RPM and translations. (a) Spins in rpm. (b) 2D translations in pixels at each frame.

feelings or just watching the sequences.

5. CONCLUSIONS

We have proposed a method for measuring the spin of a table tennis ball with Inverse Compositional Image Alignment under some assumptions that are practical for this application. The prototype system of the proposed method is useful and currently used for training of players. Although in this paper we have focused specifically on table tennis, the proposed method has a large variety of potential applications in other sports too: football, baseball, golf, volleyball, and so on. Directions of future work to make the system more practical include automatic detection of the center and radius of the ball, and the use of temporal information of multiple frames.

6. REFERENCES

[1] V. Lepetit, P. Fua, "Monocular Model-Based 3D Tracking of Rigid Objects: A Survey", *Foundations and Trends in Computer Graphics and Vision*, Now Publishers, 1(1), 1–89, 2005.

[2] Johan Philip, "Estimation Three-Dimensional Motion of Rigid Objects from Noisy Observations," *PAMI*, 13(1) 61–66, 1991.

[3] Wu Huan Qun, Qin Zhifeng, Xu Shaofa, Xu Enting, "Experimental Research in Table Tennis Spin," *International Journal of Table Tennis Sciences*, The ITTF, 1, 73–78, 1992.

[4] Z. Xiaopeng, "An experimental investigation into the influence of the speed and spin by balls of different diameters and weights", *Science and Racket Sports II*, E & FN Spon, 206–210, 1998.

[5] Vincenzo Caglioti, Alessandro Giusti, "Recovering ball motion from a single motion-blurred image," *CVIU*, 113(5), 590–597, 2009.

[6] Giacomo Boracchi, Vincenzo Caglioti, Alessandro Giusti. "Estimation of 3D instantaneous motion of a ball from a single motion-blurred image," *VISIGRAPP2009*, 225-237, 2009.

[7] H. Shum, T. Komura, "Tracking the translational and rotational movement of the ball using high-speed camera movies," *ICIP2005*, 3, 1084–1087, 2005.

[8] Christian Theobalt, Irene Albrecht, Jörg Haber Haber, Marcus Magnor, Hans-Peter Seidel, "Pitching a baseball: tracking high-speed motion with multi-exposure images," *SIGGRAPH2004*, 540-547, 2004.

[9] Alexander Szep, "Quantifying Rotations of Spheric Objects," *MVA2011*, 2011.

[10] Takuya Inoue, Yuko Uematsu, Hideo Saito, "Estimation of Rotational Velocity of Baseball Using High-Speed Camera Movies," IEEJ Transactions on Industry Applications, 131(4), pp. 608-615, 2011.

[11] Tomoyuki Nagami, Jun Morohoshi, Takatoshi Higuchi, Hiroki Nakata, Shigeto Naito, Kazuyuki Kanosue, "The Spin on Fastballs Thrown by Elite Baseball Pitchers," Medicine & Science in Sports & Exercise, 2011.

[12] Naoki Okada, Tomohiro Yabuuchi, Takuya Funatomi, Koh Kakusho, Michihiko Minoh, "Estimating Unobserved Motion of a Ball by Acquiring Its Physical Model from Image Sequences," IEICE-D, 91(12), 2950-2960, 2008.

[13] Toru Tamaki, Takahiko Sugino, Masanobu Yamamoto, "Measuring Ball Spin by Image Registration," *FCV2004*, 269–274, 2004.

[14] Bruce D. Lucas, Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *IJCAI81*, 674–679, 1981.

[15] Simon Baker, Iain Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *IJCV*, 56(3), 221-255, 2004.

[16] Simon Baker, Raju Patil, Kong Man Cheung, Iain Matthews, "Lucas-Kanade 20 Years On: Part 5," *Tech. Report CMU-RI-TR-04-64*, Robotics Institute, Carnegie Mellon University, 2004.

[17] Iain Matthews, Simon Baker, "Active Appearance Models Revisited," *IJCV*, 60(2) 135–164, 2004.