Badminton Video Analysis based on Spatiotemporal and Stroke Features

Wei-Ta Chu
National Chung Cheng University
Chiayi, Taiwan
wtchu@ccu.edu.tw

Samuel Situmeang
National Chung Cheng University
Chiayi, Taiwan
samuel.situmeang@ymail.com

ABSTRACT

Most of the broadcasted sports events nowadays present game statistics to the viewers which can be used to design the gameplay strategy, improve player's performance, or improve accessing the point of interest of a sport game. However, few studies have been proposed for broadcasted badminton videos. In this paper, we integrate several visual analysis techniques to detect the court, detect players, classify strokes, and classify the player's strategy. Based on visual analysis, we can get some insights about the common strategy of a certain player. We evaluate performance of stroke classification, strategy classification, and show game statistics based on classification results.

CCS CONCEPTS

•Computing methodologies → Activity recognition and understanding; •Human-centered computing → *Visual analytics;*

KEYWORDS

Badminton video, stroke classification, strategy classification, game statistics

ACM Reference format:
Wei-Ta Chu and Samuel Situmeang. 2017. Badminton Video Analysis based on Spatiotemporal
and Stroke Features. In *Proceedings of ICMR '17, June 6–9, 2017, Bucharest, Romania, ,* 4 pages.
DOI: http://dx.doi.org/10.1145/3078971.3079032

1 INTRODUCTION

Game statistics are important to show details of sports events. In tennis, games statistics consist of each player's number of winners, unforced errors, aces, and so on. They help coaches to design the strategy to improve their player's performance. For viewers, game statistics would improve accessing the point of interest of a game. However, research on badminton videos specifically has been rare, and most past researches either concentrates on tennis videos or generalize it to racket sports [17][1][14]. Badminton is the fastest racket game [13], and this makes playing pattern and speed different from other racket games.

Focusing on tennis videos, Reno et al. [17] proposed a background subtraction algorithm, which facilitates accurate player and ball segmentation. Barnett and Clarke [2] proposed to predict outcomes of tennis matches. They showed how the standard statistics published by the The Association of Tennis Professionals can be combined to predict the serving statistics. Polk et al. [14] presented a tennis visualization system to show detailed information of a match in an organized way. Rea et al. [16] built event hidden Markov models with binary classification according to the player's positions in the court.

Relatively fewer badminton-related studies have been proposed. Chen and Wang [5] proposed a method based on 2-D seriate images to discover statistics of a badminton match. Careelmont [3] conducted badminton shot classification in compressed videos. The shuttlecock was detected and the stroke type was recognized based on the shuttlecock trajectory. Dierickx [7] continued Careelmont's work and improved performance of the trajectory extractor significantly. Overall, these works focus on shuttlecock trajectory extraction in order to facilitate classification.

We see most badminton-related works were limited to stroke classification. In this paper, we try to integrate visual analysis techniques to detect the court, detect players, classify strokes, and classify the player's strategy. Based on visual analysis, we attempt to get insights about the common strategy of a certain player.

The rest of this paper is organized as follows. Section 2 describes details of court detection. Section 3 introduces main visual analysis techniques. Section 4 provides evaluation results, followed by the conclusion and future works in Section 5.

2 COURT DETECTION

We follow the framework proposed in [9] to detect the court. This method was originally designed to find tennis court, and we slightly modify some components to detect the court in badminton videos.

Figure 1 shows the court detection framework. Because in most cases the court lines are white, the first step of this framework is white pixel extraction. Intuitively, pixels with intensity values higher than a predefined threshold are detected as white pixels. However, many other objects such as advertisement logos, net, spectators, and the player's clothes can also be white. We therefore devise a texture filtering process as the second step to remove noise.

The idea to remove noise is to exclude white pixels that are not at lines or corners. Motivated by the corner and edge principal values proposed in [10], we consider gradient information of a local patch W centered by each white pixel located at (x, y), and construct a structure matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \sum_{u \in W} \sum_{v \in W} w(u, v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (1)$$

ICMR '17, June 6–9, 2017, Bucharest, Romania
 ACM ISBN 978-1-4503-4701-3/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3078971.3079032

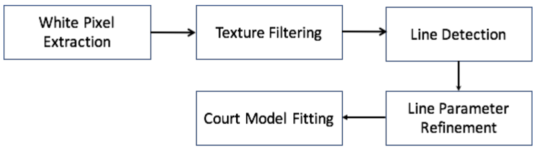


Figure 1: Overview of the court detection framework.

where $w(u, v)$ is the weighting function, and I_x and I_y are the partial derivatives in the horizontal and vertical aspects, respectively.

We then calculate the eigenvalues λ_1 and λ_2 of the structure matrix $\boldsymbol{A}$. Assuming that $\lambda_1 \leq \lambda_2$, based on the numerical values of the eigenvalues, the following observations can be made:

- If $\lambda_1 \approx 0$ and $\lambda_2 \approx 0$, then the white pixel at (x, y) has no feature of interest and should be excluded in the following process.
- If $\lambda_1 \approx 0$ and λ_2 is large, then the white pixel at (x, y) is probably on an edge and is retained.
- If both λ_1 and λ_2 are large, then the white pixel at (x, y) should be right on a corner and is retained.

Based on filtered white pixels, the Hough transform was used to detect court lines. In our work, we adopt a more recent probabilistic Hough transform [12] (PHT) to obtain more reliable results. In contrast to the standard Hough transform that takes all candidate pixels into account, PHT takes only a random subset of them to avoid noise but remain the ability to detect lines.

Results of the Hough transform or PHT may be a bundle of detected lines, which all lie close together. We thus adopt the line parameter refinement process in [9] to remove duplicate lines.

Finally, the final step shown in Figure 1 is court model fitting. At this step we need to find the mapping between the detected lines on video frames and the predefined court model. To do this, four different intersection points between court lines are extracted, and the the planar homography matrix is estimated to do projective mapping from the court in video frames to the predefined court. Details of homographic mapping please refer to [9].

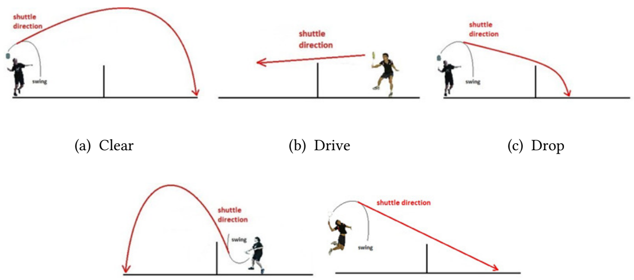
With the homography matrix, we can calibrate different video frames and represent player's positions in a standardized coordinate system, which is important when we try to develop a computational model to do strategy analysis.

3 BADMINTON VIDEO ANALYSIS

We develop three important components to facilitate getting insights from game statistics. First, because how players move conveys rich cues to understand the game, we detect player's position and map it on the court model. Second, how a player hits the shuttlecock directly show the player's performance and influence game results. We therefore do stroke classification based on player's posture. Finally, we jointly consider player's movement and stroke sequences to do strategy classification.

3.1 Player Detection and Tracking

Because players may wear clothes in various color, it is not plausible to construct a generic model to describe players in various games. Instead, we construct a background model describing the color distribution of the court. With this model, the probability of each pixel belonging to the background is estimated. The pixels with low probability being the background are viewed to present players or court lines. Obviously we can easily eliminate court line pixels thereafter because we have already detected court lines (Sec. 2).



(a) Clear (b) Drive (c) Drop

(d) Lob (e) Smash

Figure 2: Illustration of five different strokes [11].

Although the color of a badminton court looks smooth and regular, the lighting condition, object movement, and many other factors dynamically change the color distribution. The key issue is thus how to create a background model and update the model parameters dynamically to adapt to background changes, including the change of scene illumination and scene composition. Inspired by [18], we construct a Gaussian mixture model (GMM) to describe the background. Given a video sequence $f_1, f_2, ..., f_N$, where f_i denotes the ith video frame, we first describe the color distribution of the region inside the court by a GMM $\mathcal{M}$, based on the observations of the first M frames, i.e., $f_1, ..., f_M$. Every time when a new frame is coming, say f_{M+1}, we update the parameters of the GMM $\mathcal{M}$ to adapt to the latest color distribution [19]. In this work, the number of frames to be considered in background model construction is 500, i.e., $M = 500$. We use three Gaussian mixtures to construct the model, and the standard expectation-maximum algorithm is used to find parameters.

For each pixel j of a video frame f_i, we estimate the probability of this pixel being the background as $p(j|\mathcal{M})$. If the probability $p(j|\mathcal{M})$ is less than a predefined threshold δ, we view the pixel j as a candidate foreground pixel. Based on detected candidate foreground pixels, we conduct morphological operations, i.e., dilation and erosion, to further remove noise. Connected components are then determined based on the remaining foreground pixels, and finally we detect the largest connected component (CC) as the player. Note that we respectively detect the largest CC in the top half and the bottom half of a video frame, which correspond to the player far from the camera and close to the camera.

To determine the player's position, we find the bounding box of the largest connected component. The middle point of the bottom border of this box is then defined as the player's position. With the homography matrix determined in Sec. 2, we can map the player's position from video to the court model.

3.2 Stroke Classification

There are several types of strokes in badminton games, and a high level of skill is required by a player to perform all of them effectively.

In this work, we generalize the strokes to six types, namely *clear*, *drive*, *drop*, *lob*, *smash*, and all other strokes are viewed as *others*. Figure 2 illustrates the first five strokes.

- Clear: Shuttle travels high and deep.
- Drive: Shuttle travels fast and flat skimming the net.
- Drop: Shuttle travels downwards.
- Lob: The aim is to lift or lob the shuttle over the opponent and to make the shuttle land closely to the baseline.
- Smash: Shuttle travels at speed in downward direction.
- Others: Other types of strokes or player standing without hitting the shuttle.

The idea to classify the stroke a player invokes is based on pose recognition. Based on the bounding box of a player, we extract histogram of oriented gradient (HOG) [6] as the representation. Based on HOG, we construct a classifier based on support vector machine (SVM) [4] to do stroke classification.

3.3 Strategy Classification

Strategy refers to the general plan that a player decides on to defeat the opponent, on the basis of some specific tactics [8]. Currently, we roughly classify strategies in badminton into either offensive or defensive strategies. Given a play, we detect player's position and classify strokes for each frame. A sequence of frames is then represented by a sequence of observations showing the evolution of player's position and strokes. More particularly, the observations are stroke type s and the bottom player's position in the court model (x, y). Note that we divide the court (bottom half) into nine equally-sized regions, and the player's position is spatially quantized into one of the nine regions. The representation of player's position is then $r = Q(x, y)$, where Q is the quantization function. A play consisting of T frames is then represented as $\boldsymbol{O} = (\boldsymbol{o}_1, ..., \boldsymbol{o}_T)$, where $\boldsymbol{o}_i = (s_i, r_i)$ is a tuple showing the stroke type and position.

Given a set of observation sequences $\{\boldsymbol{O}_1, ..., \boldsymbol{O}_K\}$, where K is the total number of training sequences, we construct hidden Markov models (HMMs) to do strategy classification. This model is trained based on the standard Baum-Welch algorithm [15]. We respectively collect training data of the offensive strategy and the defensive strategy, and separately train a model for each strategy. Given a play, we respectively input the observation sequence $\boldsymbol{O}$ to these two models λ_1 and λ_2. The probability of this play being an offensive strategy is estimated as $p(\boldsymbol{O}|\lambda_1)$, and the probability of being a defensive strategy is estimated as $p(\boldsymbol{O}|\lambda_2)$. The strategy of this play is finally determined as $i^* = \arg\max_i p(\boldsymbol{O}|\lambda_i)$.

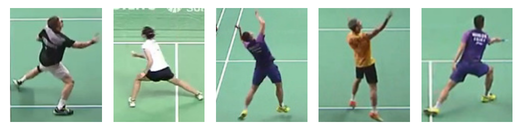
4 EVALUATION

4.1 Dataset and Evaluation Setup

We collect the evaluation dataset from BadmintonWorld.TV channel in YouTube. In this work, we mainly focus on the match from top-down camera view, and eliminate highlight replay. Each video is divided into two or four parts according to the number of sets in the match. The third set is divided into two parts because the players must switch their position when a player's score reaches 11 points. Length of each set differs from two to eleven minutes, and the average length of a play is around 10 seconds. Resolution of each video is 854×480. There are 6 videos covering 558 plays in total. Table 1 shows detailed information of the dataset.

Table 1: Detailed information the evaluation dataset.

Match	Sets		
	1	2	3
TOTAL BWF World Championships 2015 Badminton Day 7 F M3-MS Chen vs Lee	32 plays	36 plays	
Victor Korea Open 2016 Badminton F M4-MS Son Wan Ho vs Qiao Bin	32 plays	44 plays	27 plays
Victor Korea Open 2016 Badminton SF M3-MS Wong Wing Ki Vincent vs Qiao Bin	38 plays	37 plays	
Yonex Denmark Open 2016 Badminton F M5-WS Akane Yamaguchi vs Tai Tzu Ying	36 plays	34 plays	30 plays
Yonex Denmark Open 2016 Badminton SF M4-WS Carolina Marin vs Akane Yamaguchi	32 plays	38 plays	38 plays
BCA Indonesia Open 2016 Badminton F M3-MS Jan O Jorgensen vs Lee Chong Wei	32 plays	37 plays	35 plays



(a) Clear (b) Drive (c) Smash (d) Smash → Clear (e) Lob → Drive

Figure 3: Sample images of correct stroke classification (left three) and false classification (right two).

To evaluate stroke classification, 600 images are manually extracted and cropped from badminton videos. In this work, we only do stroke classification for the bottom player. For each stroke type we collect 100 images. The resolution of these images is 120×150.

4.2 Performance of Player Detection

Denote the region of the detected player as D and the region of the ground truth G, we say the player is correctly detected if $\frac{D \bigcap G}{D \bigcup G} > 0.5$. Overall, we obtain 85.45% accuracy in player detection. One problem of the proposed player detection module is that shadow of the player may be misclassified as a part of the player, and disturbs player's positions.

4.3 Performance of Stroke Classification

We adopt the five-fold cross validation scheme to evaluate stroke classification. At each fold, we use 480 stroke images for training and 120 stroke images for testing. The number of correctly classified strokes divided by 120 is calculated as the classification accuracy. Overall, we obtain 83.33% average accuracy.

Figure 3 shows sample images of correct classification and false classification. The left three strokes are correctly classified into clear, drive, and smash, respectively. Figure 3(d) is actually a smash stroke, but is misclassified into clear. By comparing Figure 3(d) with Figure 3(a), we see the similarity between these two poses. Figure 3(e) is actually a lob, but is misclassified into drive. We again observe the similarity between Figure 3(e) and Figure 3(b). To achieve more accurate classification, we will further consider motion information in the future.

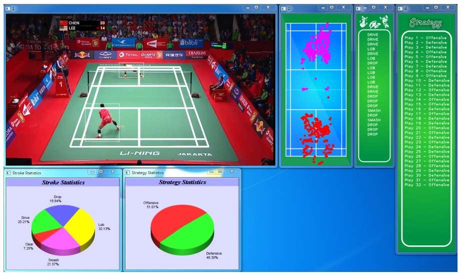


Figure 4: The interface to show game statistics. The windows from left to right, top to bottom, show the original video, player's position evolution, stroke classification results, strategy classification results, distribution of strokes, and distribution of strategies.

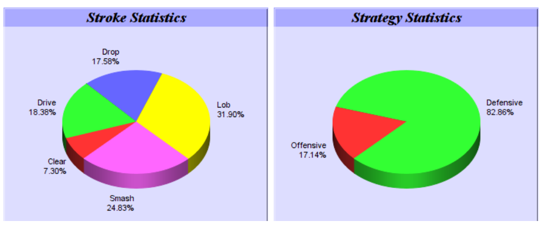


Figure 5: Sample distributions of strokes (left) and strategies (right).

4.4 Performance of Strategy Classification

We also use the five-fold cross validation scheme to evaluate strategy classification. At each fold, we use 18 plays for training and 6 plays for testing. Based on automatically extracted spatiotemporal features, the average strategy classification accuracy is 70%. Currently we don't work perfectly in stroke classification and player position detection. If the strategy classification model is constructed based on manually-corrected spatiotemporal features and stroke classification results, this system can achieve 83.33% average accuracy in strategy classification.

4.5 Badminton Video Statistics

Based on classification results of stroke classification, player detection, and strategy classification, we can see details of a badminton match, and provide more cues and insights by visualization. Figure 4 shows the interface showing various classification and detection results. From left to right, top to bottom, the sub-windows in Figure 4 are the original video, continuous player's positions illustrated in the court model, stroke classification results, strategy classification results, stroke distribution, and strategy distribution.

To get more insights, we can go into the details of each set. Figure 5 shows the stroke distribution and the strategy distribution of the game "TOTAL BWF World Championships 2015 Chen vs Lee". In this set, we observe Lee's stroke and strategy statistics. Lee used more lob strokes than other strokes. Statistics of the strategy shows that Lee focuses on defensive strategies. In fact, he was forced to do this, because his opponent attacked quite a lot in this set. Lee lost this set in the end.

5 CONCLUSION

In this paper, we integrate several visual analysis techniques to detect the court, detect the players, classify strokes, and classify player's strategy. We propose an adaptive background modeling to facilitate player detection. Based on player's pose, we classify the stroke a player invokes. By modeling the evolution of player's position and stroke information, we recognize a player's strategy. With these results, we visualize information of a badminton match, which can be used to analyze the player's performance and understand the characteristics of this match.

In the future, more elegant strategy modeling and more spatiotemporal features are to be pursued. In addition, motion information will be considered in stroke classification and strategy classification.

ACKNOWLEDGMENTS

The work was partially supported by the Ministry of Science and Technology of Taiwan under the grant MOST105-2628-E-194-001-MY2, MOST104-2221-E-194-014, and MOST103-2221-E-194-027-MY3.

REFERENCES

[1] T. Asano, Y. Serikawa, and S. Itoh. 2015. Quantitative Analysis of Tennis Ball Rotation and Trajectory. In *Proceedings of IEEE International Symposium on Industrial Electronics*. 906–911.

[2] T. Barnett and S.R. Clarke. 2005. Combining Player Statistics to Predict Outcomes of Tennis Matches. *IMA Journal of Management Mathematics* 16, 2 (2005), 113–120.

[3] S. Careelmont. 2013. *Badminton Shot Classification in Compressed Video with Baseline Angled Camera*. Master Thesis, University of Ghent.

[4] C.-C. Chang and C.-J. Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3 (2011), 1–27.

[5] B. Chen and Z. Wang. 2007. A Statistical Method for Analysis of Technical Data of a Badminton Match based on 2-d Seriate Images. *Tsinghua Science and Technology* 12, 5 (2007), 594–601.

[6] N. Dalal and B. Triggs. 2005. Histograms of Oriented Gradients for Human Detection. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. 886–893.

[7] T. Dierickx. 2014. *Badminton Game Analysis from Video Sequences*. Master Thesis, University of Ghent.

[8] J. Downey. 2007. *Tactics in Badminton Singles*. http://www.badminton-information.com/jake-downey-biography-badminton-books.html.

[9] D. Farin, S. Krabbe, W. Effelsberg, and P.H.N. de With. 2004. Robust Camera Calibration for Sport Videos using Court Models. In *Proceedings of SPIE Storage and Retrieval Methods and Applications for Multimedia*, Vol. 5307. 80–91.

[10] C. Harris and M. Stephens. 1988. A Combined Corner and Edge Detector. In *Proceedings of Fourth Alvey Vision Conference*. 147–151.

[11] Master Badminton 2017. *Badminton Techniques, Shots and Skills*. Master Badminton. http://www.masterbadminton.com/badminton-techniques.html.

[12] J. Matas, C. Galambos, and J. Kittler. 2000. Robust Detection of Lines Using the Progressive Probabilistic Hough Transform. *Computer Vision and Image Understanding* 78, 1 (2000), 119–137.

[13] Oceania Badminton Confederation 2011. *World's Fastest Racquet Sport*. Oceania Badminton Confederation. http://websites.sportstg.com/assoc_page.cgi?c=1-1065-0-0-0&sID=13337&&news_task=DETAIL&articleID=5850561.

[14] T. Polk, J. Yang, Y. Hu, and Y. Zhao. 2014. TenniVis: Visualization for Tennis Match Analysis. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2339–2348.

[15] L. Rabiner. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* 77, 2 (1989), 257–286.

[16] N. Rea, R. Dahyot, and A. Kokaram. 2005. Classification and Representation of Semantic Content in Broadcast Tennis Videos. In *Proceedings of IEEE International Conference on Image Processing*. 1204–1207.

[17] V. Reno, N. Mosca, M. Nitti, T. Dorazio, D. Campagnoli, A. Prati, and E. Stella. 2015. Tennis Player Segmentation for Semantic Behavior Analysis. In *Proceedings of IEEE International Conference on Computer Vision Workshop*. 718–725.

[18] Z. Zivkovic. 2004. Improved Adaptive Gaussian Mixture Model for Background Subtraction. In *Proceedings of International Conference on Pattern Recognition*. 28–31.

[19] Z. Zivkovic and F. van der Heijden. 2006. Efficient Adaptive Density Estimation per Image Pixel for The Task of Background Subtraction. *Pattern Recognition Letters* 27, 7 (2006), 773–780.