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:
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. 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.

We see most 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.

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 centered by each white pixel located at (x, y), and construct a structure matrix A:

$$A = \sum_{u \in W} \sum_{v \in W} w(u, v) \begin{bmatrix} I_u^2 & I_u I_v \\ I_v I_u & I_v^2 \end{bmatrix}$$

(1)
We develop three important components to facilitate getting in-
where we detect player’s position distribution of the court. With this model, the probability of each
Instead, we construct a background model describing the color
to construct a generic model to describe players in various games.
Because players may wear clothes in various color, it is not plausible
3.1 Player Detection and Tracking
sequences to do strategy classiﬁcation.
posture. Finally, we jointly consider player’s movement and stroke
results. We therefore do stroke classiﬁcation based on player’s

Figure 1: Overview of the court detection framework.

where \( w(u, v) \) is the weighting function, and \( l_x \) and \( l_y \) are the partial
derivatives in the horizontal and vertical aspects, respectively.
We then calculate the eigenvalues \( \lambda_1 \) and \( \lambda_2 \) of the structure
matrix \( A \). 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 \( \lambda_2 \) is large, then the white pixel at \((x, y)\) is
  probably on an edge and is retained.
- If both \( \lambda_1 \) and \( \lambda_2 \) 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 in-
sights from game statistics. First, because how players move con-
vveys 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).

Although the color of a badminton court looks smooth and reg-
ular, 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 \( f_1, f_2, ..., f_M \), where \( f_i \)
denotes the ith video frame, we ﬁrst describe the color distribution
of the region inside the court by a GMM \( M \), based on the observations
of the ﬁrst M frames, i.e., \( f_1, f_2, ..., f_M \). Every time when a new frame
is coming, say \( f_{M+1} \), we update the parameters of the GMM \( M \)
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 \( p(j|M) \). If the probability
\( p(j|M) \) is less than a predeﬁned threshold \( \delta \), we view the pixel
as a candidate foreground pixel. Based on detected candidate fore-
ground 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 fi-
nally 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 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 \( O = (o_1, ..., o_T) \), where \( o_i = (s_i, r_i) \) is a tuple showing the stroke type and position.

Given a set of observation sequences \( \{O_1, ..., O_K\} \), 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.

We adopt the five-fold cross validation scheme 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 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>
<thead>
<tr>
<th>Match</th>
<th>Sets</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL BWF World Championships 2015 Badminton Day 7 F M3-MS Chen vs Lee</td>
<td>32 plays</td>
<td>36 plays</td>
<td>36 plays</td>
<td></td>
</tr>
<tr>
<td>Victor Korea Open 2016 Badminton F M4-MS Son Wan Ho vs Qiao Bin</td>
<td>32 plays</td>
<td>44 plays</td>
<td>27 plays</td>
<td></td>
</tr>
<tr>
<td>Victor Korea Open 2016 Badminton SF M3-MS Wong Wing Ki Vincent vs Qiao Bin</td>
<td>38 plays</td>
<td>37 plays</td>
<td>37 plays</td>
<td></td>
</tr>
<tr>
<td>Yonex Denmark Open 2016 Badminton M5-WS Akane Yamaguchi vs Tai Tzu Ying</td>
<td>36 plays</td>
<td>34 plays</td>
<td>30 plays</td>
<td></td>
</tr>
<tr>
<td>Yonex Denmark Open 2016 Badminton SF M4-WS Carolina Marin vs Akane Yamaguchi</td>
<td>32 plays</td>
<td>38 plays</td>
<td>38 plays</td>
<td></td>
</tr>
<tr>
<td>BCA Indonesia Open 2016 Badminton F M3-MS Jan O Jorgensen vs Lee Chong Wei</td>
<td>32 plays</td>
<td>37 plays</td>
<td>35 plays</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 3](image)

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

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 \( D \cap G \neq \emptyset \). 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 85.33% 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.
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.

5 CONCLUSION

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