Detection of Shot Information Using Footwork Trajectory and Skeletal Information of Badminton Players

Naoki Tanaka[1], Hidehiko Shishido[2][a], Masashi Suita[3], Takeshi Nishijima[4][b], Yoshinari Kameda[2][c] and Itaru Kitahara[2][d]

[1]*Master's and Doctoral Program in Intelligent Mechanical Interaction Systems, University of Tsukuba, Ibaraki, Japan*
[2]*Center for Computational Sciences, University of Tsukuba, Ibaraki, Japan*
[3]*Faculty of Health and Sport Sciences, University of Tsukuba, Ibaraki, Japan*
[4]*Faculty of Health Sciences, Tokyo Metropolitan University, Tokyo, Japan*

Keywords: Shot Information, Footwork Trajectory, Skeletal Information, Video Analysis, Badminton.

Abstract: As video analysis has become important for sports science, various research has been conducted. In badminton, while shot information is essential primary data for performance analysis, it has been input manually, which makes it difficult to give instant feedback onsite. Our research aims to automatically detect shot information from videos of badminton game. By applying video tracking, the player's footwork trajectory and skeletal information are estimated. Based on the estimated information, the hit timing is detected using deep learning classification. The horizontal position of the hit point, which is useful for game analysis, is also detected from the player's footwork trajectory around the hit timing.

1 INTRODUCTION

With the development of IT, data analytics is being considered fundamental for sports science, especially in competitive sports (Pascual, 2006), (Messelodi, 2019). Since analysis using video/image data is particularly noteworthy as information can be confirmed visually, there are active research to analyze players' body information in badminton competition (Kokum, 2017), (David, 2014) and shot information (Chen, 2007), (Nyan, 2020), (Yoshikawa, 2021). These research detect players' forms and tactics to improve their competitive performance.

A badminton game consists of multiple rallies in which players exchange shots, and each rally includes a variety of shots. Thus, the shot information is considered as the primary data for analysing the evolution of rallies that lead to scored and lost. On this demand, analytical software such as "Sportscode" is widely used for video analysis of badminton shot information. It can analyze player performance by creating a database of competition scenes. For data aggregation, the timing of shots taken by players and patterns of goals scored is manually input. On the other hand, badminton matches take more than one hour, and collecting data using manual input requires lots of time and effort.

In order to solve the problem of badminton game analysis, as shown in Figure 1, this research proposes a method for automatically detecting shot information from monocular badminton singles match videos. The detected shot information consists of two properties of the hit point: "Hit timing" and "Horizontal position".

- [Hit timing] The time when the player hits a shuttle.
- [Horizontal position] Vertical projection of the 3D position where a racket hits a shuttle onto the court (ground).

[a] https://orcid.org/0000-0001-8575-0617
[b] https://orcid.org/0000-0002-6765-4282
[c] https://orcid.org/0000-0001-6776-1267
[d] https://orcid.org/0000-0002-5186-789X

Tanaka, N., Shishido, H., Suita, M., Nishijima, T., Kameda, Y. and Kitahara, I.
Detection of Shot Information Using Footwork Trajectory and Skeletal Information of Badminton Players.
DOI: 10.5220/0012162700003587
In *Proceedings of the 11th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2023)*, pages 112-119
ISBN: 978-989-758-673-6; ISSN: 2184-3201

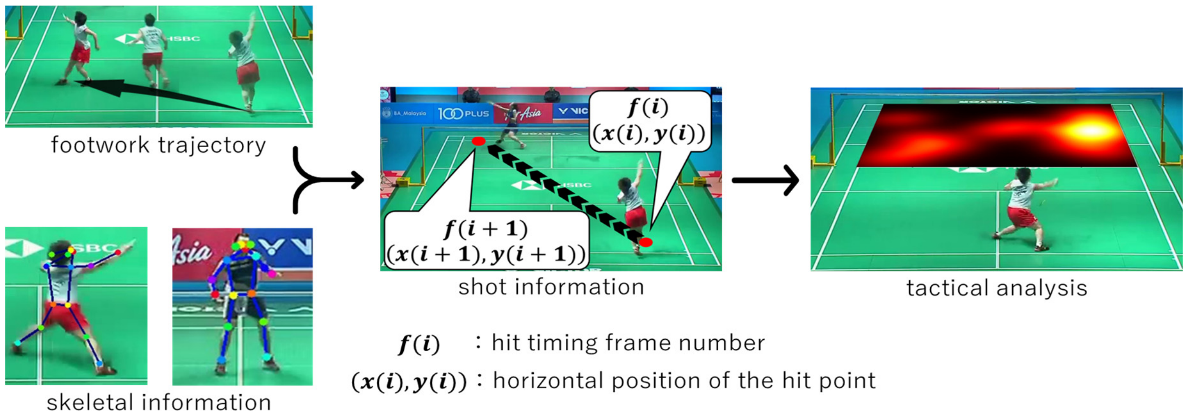

Figure 1: Shot information detection and tactical analysis using footwork trajectory and skeletal information. The shot information is detected using footwork trajectory (upper-left) and skeletal information (lower-left) from a monocular video of a badminton game. The detected shot information consists of the hit timing and the horizontal position of each hit point (center). Tactics are analyzed by visualizing the ratio of shot direction based on the detected shot information as shown in the right.

In this research, it is assumed that the positions of the players and their dominant wrists show time-series changes correlation with the hit timing. Since these time-series changes are more pronounced in singles matches, this research focuses on singles matches. Further contributes of this research is to solve the data collection problem by evaluating tactical analysis using the detected data.

2 RELATED WORKS

2.1 Human Pose Estimation

Some 3D tracking sensors such as "Vicon" and "OptiTrack" require attaching markers to the body for estimating a subject's pose. Although, this kind of method has high accuracy in pose estimation (the skeletal position with an error of 0.1 mm or less), applying this method to a live sporting event is difficult, due to some limitation to the players' movement,

Some methods can estimate a subject's pose from images and videos (Zhe, 2019), (Sun, 2019). There are two major approaches to this method: top-down and bottom-up. The top-down method first detects subjects using object detection and then detects joints for each person. On the other hand, the bottom-up method detects all joint points in an image without object detection and then compiles the joint points constituting each subject. While the top-down method is generally more accurate, the processing time increases significantly when the number of subjects increases. On the other hand, the disadvantage mentioned above is less likely to occur due to the small and fixed number of players in the badminton game video used in this research. Therefore, a top-down person poses estimation method is used in this research.

Human pose estimation from images has been applied to various kind of sports. For example, Dejan et al. proposed a system to estimate the pose of a jumper and a skiboard in ski jumping videos and evaluate the pose of the jumper during jumping (Dejan, 2022). Nonaka et al. proposed a system to estimate the pose of a player in a tackle scene in a rugby video and to judge dangerous tackles that may cause concussions (Nonaka, 2022). Kaustubh et al. estimated players' poses in table tennis videos (Kaustubh, 2021). They classified hitting styles using deep learning with the time-series information as input, achieving more than 99% accuracy in 11 hitting style classification.

2.2 Badminton Shot Information Analysis

Shot information is helpful for tactical analysis of badminton games. Wei et al. classified shots into six types (five plus others) based on players' pose information at hit timing (Wei, 2017). Wei et al. also analyzed the game situation based on the usage of each shot. The analysis results were compared with the actual match results. It was confirmed that the results of the game situation analysis obtained from the shot information were equivalent to the actual match results. In this research, we attempted to detect shot information in consideration of its applicability to the tactical analysis of badminton.

2.3 Detection of Shot Information

Shuttle tracking has been reported as one of the most common shot information detection methods. By tracking the shuttle, information on the shuttle's position and trajectory can be obtained and information on each shot can be detected. There is an example of detecting the shuttle in an image using the image subtraction method and comparing the shuttle's trajectory with various strokes (Chen, 2007). In another case (Nyan, 2020), the authors applied Tracknet and polynomial curve fitting to the shuttle's trajectory using deep learning and attempted to detect shot information. All these research assume that the shuttle is always observed within the angle of view. However, since badminton competition video is generally shot from a single fixed camera focusing on the players, there are frames in which the shuttle is not seen in the video. These frames may make it difficult to track the shuttle.

Therefore, a method to detect shot information without tracking shuttles was proposed (Yoshikawa, 2021). The hit timing is identified from the player's skeletal information. The shot direction is detected by connecting the hit points of the two players at the hit timing in a time series. However, since there are countless possible forms in badminton depending on the shuttle's position, more than classification based on a small number of conditional branches is required to identify the hit timing for all forms.

This research aims to detect shot information using players' footwork trajectories and skeletal information. By achieving this objective, the following two contributions can be obtained.

- The detection of shot information that can be applied to match videos, including scenes in which the shuttle is not visible.
- The detection of shot information that can be applied to various shots.

3 PRELIMINARY EXPERIMENTS

Our research uses a player's 2D skeletal information (16 - 25 joint points) from monocular videos and shot information detection. However, there are some problems with inputting all the acquired skeletal information into the deep learning model as follows.

- The accuracy of shot information detection is highly dependent on the accuracy of the conventional human skeleton estimation method.
- Increased computational complexity and over-fitting risk due to many input dimensions.

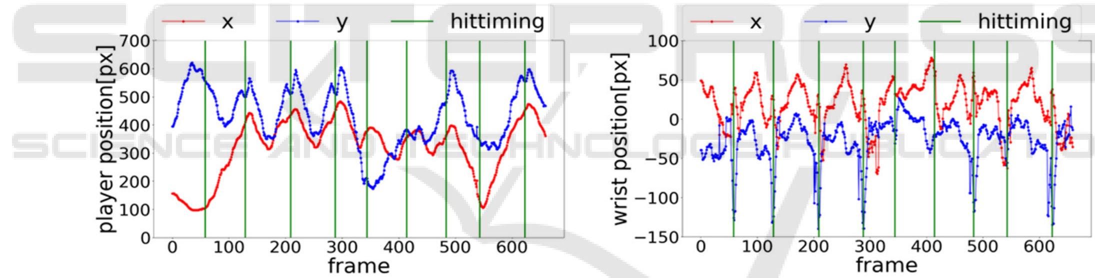

Figure 2: Relationship between time-series changes in player position (upper) and wrist position (lower), respectively, and hit timing.

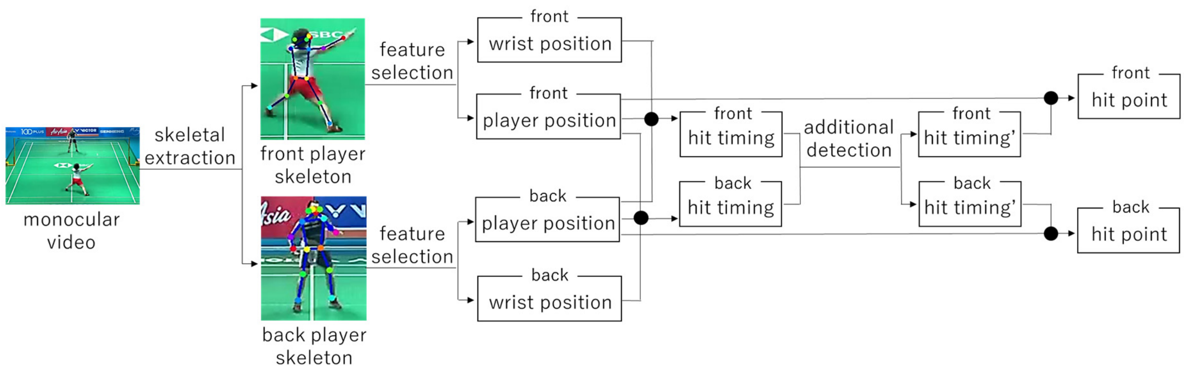

Figure 3: Overview of the shot information detection method. The skeletal information of the two players is detected from the monocular video, and the player's position and dominant wrist position are extracted. These are input to LSTM to detect the hit timing and the ground projection of the hitting point. “additional detection” is explained in Section 4.2.

Thus, in this research, only the player's position (the midpoints of both ankles) and wrist position (the relative position of the dominant wrist from the waist) are used among the acquired skeletal information to address the abovementioned problem. These features were selected in consideration of the features of badminton games, and their validity as features were verified by visualizing the relationship with the hit timing.

In badminton singles, the base is usually located near the center of the court. The player has the feature of moving back to the base after his shot to prepare for the opponent's hit. Therefore, as shown in Figure 2 (left), the time-series variation of the player's position (x, y) is correlated with the hit timing. Furthermore, since the player's position at the hit timing is the position where the player moved to release the shot. Thus, the player's position around the hit timing and the horizontal position of the hit point are close to each other.

The player's dominant hand holding the racket has the postural features of being the most distant from the body center at the hit timing. Therefore, as shown in Figure 2 (right), the relative position of the dominant wrist from the waist (x, y) also correlates with the hit timing. Therefore, this information has features time-series changes at the hit timing and practical features for detecting the hit timing and the horizontal position of the hit point.

4 DETECTION OF SHOT INFORMATION

As shown in Figure 3, this research proposes a method for detecting shot information, such as hit timing and the horizontal position of the hit point. A deep learning based approach is taken where the network takes the physical features at the hit timing as the input. As shown in Figure 4, the conventional method of human skeleton estimation (Sun, 2019) is applied to a badminton competition video to obtain the 2D skeletal information of the player. The player's and wrist positions are calculated based on the skeletal information. Hit timing is detected from the features of the time series data of the player's and wrist positions. The horizontal position of the hit point can be detected from the player's footwork trajectory centered on the detected hit timing.

4.1 Detection of Hit Timing

This section describes the preprocessing and output data for the input data of the hit timing (as shown in Figure 1: ***f(i)***) detection model.

As shown in Figure 5, the elements used as input data for the model are the time-series data of player positions and wrist positions obtained in Section 4.1. The player position data is for two players, one who releases the shot and the other who receives the shot. The wrist position data is for one player who releases the shot. A sliding window of step s and length w is applied to these time series data to generate partial time series data. The model uses LSTM with the input partial time-series data at each time point. The model output the probability that a hit timing is included in the time of the partial input time series data.

Figure 4: Acquisition of skeletal information and player position. Bule crosses is the midpoint of both ankles and defined as the player position.

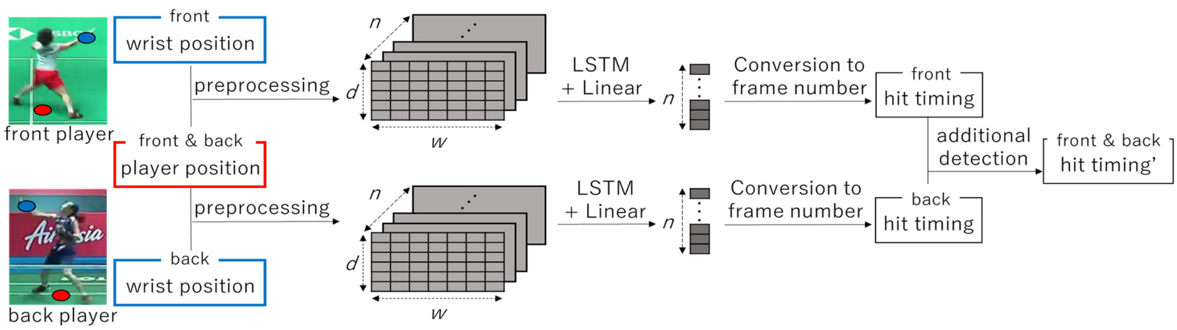

Figure 5: Process flow of hit timing detection method. A sliding window is applied to the time series data of the positions of the two players and their dominant wrists to generate partial time series data. The generated partial time-series data is input to LSTM for hit timing detection. "Conversion to frame number" and "additional detection" are explained in section 4.2. d: dimensionality of features, w: window width of a sliding window, n: number of partial time series.

Cross entropy error between the correct label and the output probability is calculated. The model is trained to make the error small.

Furthermore, the following condition is applied to the hit timing detection results of the two players to reduce the number of undetected hit timing.

- If the hit timing of one player is detected consecutively, the frame with the largest output probability for the other player between the two consecutive hit timing is additionally detected as a hit timing.

This condition is to take advantage of the competitive features that two players always alternate shots in a badminton game.

The accuracy of hit timing detection can be evaluated by comparing the final hit timing prediction frame (Figure 5: hit timing') with the ground truth after adding this additional detection.

4.2 Detection of the Horizontal Position of the Hit Point

This section describes the preprocessing for the model's input and output data for detecting the horizontal position of the hit point (as shown in Figure 1: (***x(i)***, ***y(i)***)).

As shown in Figure 6, the elements used as inputs to the model are the time series data of player position obtained in Section 4.1 and the hit timing obtained in Section 4.2. From the time series data of player positions, partial time series data of length *w* centered at each hit timing is extracted. Using LSTM with the partial time series data of each hit timing as input, the model output the 2D coordinates of the horizontal position of the hit point.

MSE (mean square error) is calculated using the ground truth as the horizontal position of the hit point. The model is trained so that the error is small.

5 EXPERIMENTS

The detection accuracy of the hit timing and the horizontal position of the hit point was verified, respectively. Furthermore, from the analysis results using the shot directions obtained from the horizontal position of the hit points of two players, the possibility of using the detected data for analysis is confirmed.

5.1 Experiment Setup

The data used in the validation are videos of two games (four sets in total) in Akane Yamaguchi vs. Tai Tzu Ying in 2019 and 2021, with each set as Video No.1 to 4. Videos No. 1 to 3 were used as training data and videos No. 4 as test data. HRNet (Sun, 2019) was used to extract the skeleton of the players. HRNet is an algorithm that estimates the skeletal shape of a person in a 2D image by deep learning and is capable of real-time detection even when multiple people are mixed in the image. The parameters in the algorithm are s=5, w=40 and df=14. *Precision*, *recall*, and *f-measure* are used as hit timing detection accuracy indices.

5.2 Experiment to Detect the Hit Timing

In this research, the detected hit timing is used to detect the horizontal position of the hit point. However, the impact of an error of a few frames on the hit timing detection is small. Because the model inputs the footwork trajectories of players for several tens of frames around the hit timing. Therefore, this research does not require strict accuracy of hit timing frame detection in units of a few frames. Considering the footwork speed of the player, a specific error (E frames) from the ground truth of the hit timing frame can be tolerated. In this experiment, E=20 (about 0.67 seconds) was set based on observing player speed in the

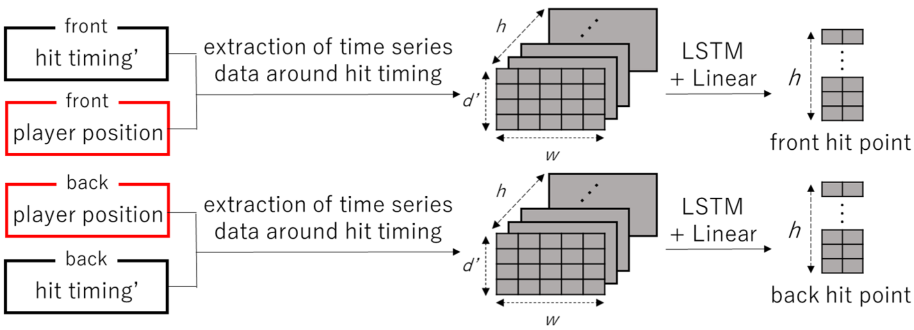

Figure 6: Processing flow of the method for detecting the horizontal position of the hit point. Extract the time series data of the player's position around the hit timing. The extracted time-series data is input to LSTM to detect the horizontal position of the hit point. d': dimensionality of features, w: window width of a sliding window, h: number of hit timing.

Table 1: Accuracy evaluation of hit timing detection.

input	recall(%)	precision(%)	f_measure(%)
pos_f	79.3	82.1	80.7
pos_b	76.0	79.3	77.6
wri_f	91.7	94.1	92.9
pos_f, pos_b	80.2	78.9	79.5
pos_f, wri_f	90.1	91.6	90.8
pos_b, wri_f	95.9	94.3	95.1
pos_f, pos_b, wri_f	93.4	**96.6**	95.0
$pos_f, pos_b, wri_f + add$	**96.7**	95.1	**95.9**

data. In other words, a detection whose error from the ground truth of the hit timing frame is within 20 frames is defined as a correct detection.

Table 1 shows the hit timing detection accuracy for the players at the front of the court. The three types of data used for hit timing detection are denoted as pos_f (position of the player at the front of the court), pos_b (position of the player at the back of the court), and wri_f (wrist position of the player at the front of the court). The detection accuracy of the model when these data are combined and input, respectively, is shown. The model with the additional detection described in section 4.2 is denoted as *+add*.

5.3 Experiments to Detect the Horizontal Position of the Hit Point

Since hit timing is used to detect the horizontal position of a hit point, the accuracy of hit timing detection affects the accuracy of the detection of the horizontal position of the hit point. In this section, the ground truth is used for the hit timing to evaluate only the detection of the horizontal position of the hit point. As shown in Figure 7, the x-axis is set parallel to the net, and the y-axis is set perpendicular to the net, with the yellow circle in the upper left corner as the origin in the overhead view of the court. Considering the half-court size of an actual badminton singles court (5.18[m]×6.70[m]), the output coordinates are set as x=[0~5.18] and y=[0~6.70]. *RMSE (root mean square error)* is used to measure detection accuracy. *RMSE* of the output and the ground truth for all test data was *RMSE*=0.54[m].

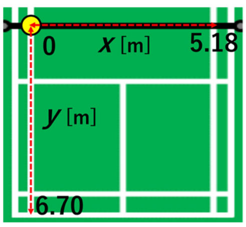

Figure 7: Coordinate system of the projected position of the hit point on the ground.

5.4 Evaluation of Analysis by Percentage of Shot Direction

This section describes the tactical analysis of players using the horizontal position of the hit points detected by this method and the possibility of applying the detected data to the analysis. As an example of analysis, this section presents an analysis method that divides a half-court into nine (3×3) areas (Josue, 2020), (Careelmont, 2013). This analysis reveals player features based on the percentage of shot directions from a specific area. Among the nine areas of the court surface, the hit area is defined as the area that includes the horizontal position of the hit point, and the shot direction can be detected by connecting the hit areas (or the horizontal position of the hit point) of two players in a time series. The ratio of shot directions can be used for tactical analysis as an indicator of the features of each player's game.

Using the correct label and detected hit timing data used in this method, the horizontal position of the hit point can be detected from the three models shown in Table 2. The G-model is a model that uses the ground truth of the horizontal position of the hit point. The GD-model is a model that uses the horizontal position of the hit point detected using the ground truth of the hit timing. The DD-model is a model that uses the horizontal position of the hit point detected using the detected hit timing. Therefore, the three models can also detect the hit area and shot direction.

The features of players can be visualized by plotting the distribution of shot direction as a heat map. Furthermore, by comparing the heat maps for each input information, the applicability to the detection data analysis can be qualitatively evaluated. Figure 8 shows the heat maps of the distribution of shot directions in the hit area where the number of hits by players in the front of the court was high in the test data. The hit timing and the horizontal position of the hit points detected by this method can lead to the same level of analysis results as those using the ground truth.

Table 2: Three models for detecting the horizontal position of the hit point.

model	hit timing	hit point
G-model		ground truth
GD-model	ground truth	detected data
DD-model	detected data	detected data

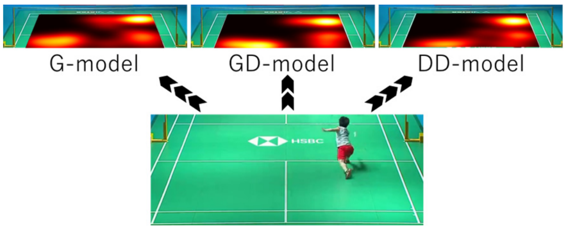

Figure 8: Visualization of shot direction percentage from the right front area of the court. The shot directions detected from the G-model (left), GD-model (center), and DD-model (right) are visualized using heat maps. The yellowish region indicates the most likely location of the endpoint of the shot direction.

5.5 Limitations in this Research

This section describes the limitations of the detection accuracy and the primary cause of error in this study.

It is difficult to detect hit timing in a single frame using this research method. This is because the output information (hit timing) is not uniquely determined from the time-series changes in the input information (movement trajectory and skeletal information). Therefore, applying this method to tactical analysis is difficult, which requires hit timing accuracy on a frame-by-frame basis.

The primary cause of error in the experiment described in Section 5.2 is the irregular movement of players in hit timing. In this study, the following hypotheses are formulated for the movement trajectory and skeletal information, respectively, as the regular movements of the players.

- The player moves away from the base around the hit timing and returns to the base again.
- At the hit timing, the player's wrist is farthest away from the waist.

Therefore, detection errors are likely to occur when players' movements that do not conform to the above hypotheses (irregular movements) are observed.

We have attempted to reduce the number of errors by applying the additional detection described in Section 4.2, but we have not been able to address all errors.

6 CONCLUSIONS

This research proposes a method for detecting shot information using badminton players' footwork trajectories and skeletal information. The proposed method detects the hit timing using the time series information of the footwork trajectory and skeletal information, which are characterized by the hit timing. Furthermore, this method detected the horizontal position of the hit point using the footwork trajectory around the hit timing. As a result of the demonstration experiment, the hit timing detection accuracy (F-measure) for the player at the front court was 95.9%, and the horizontal position of the hit point was detected with an accuracy of MSE=54.0. Using the detected data, we quantitatively and qualitatively evaluated that the tactical analysis can be performed at the same level as the ground truth. As a future work, we plan to expand the experimental data. We will confirm the generality of the method from the results of tests using players' data that are not included in the training data. We will also investigate how players' playing styles and other factors affect detection accuracy.

This method can automatically detect data for tactical analysis using only the player's footwork trajectory and skeletal information from the monocular camera images. In conclusion, this research succeeded in solving the problem of the conventional method using shuttle tracking and in automating the data collection for badminton video analysis.

REFERENCES

Pascual J. Figueroa, Neucimer J. Leite, Ricardo M. L. Barros, (2006) "Tracking soccer players aiming their kinematical motion analysis, " Computer Vision and Image Understanding Vol.101, pp.122-135.

S. Messelodi, C.M.Modena, V.Ropele, S.Marcon, M.Sgro, (2019) "A Low-Cost Computer Vision System for Real-Time Tennis Analysis, " ICIAP, pp.106-116.

Kokum Weeratunga, Anuja Dharmaratne, Khoo Boon How, (2017) "Application of computer vision and vector space model for tactical movement classification in badminton, " IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol.2017, pp.132-138.

David Alder, Paul R.Ford, Joe Causer, A.Mark Williams, (2014) "The coupling between gaze behavior and opponent kinematics during anticipation of badminton shots, " Human Movement Science, Vol.37, pp.167- 179, Oct.

Chen Bingqi, Wang Zhiqiang, (2007) "A Statistical Method for Analysis of Technical Data of a Badminton Match Based om 2-D Seriate Images, "Tsinghua Science and Technology ISSN 1007-0214 14/19 pp.594-601, Vol.12, No.5.

Nyan-Ping Ju, Dung-Ru Yu, Tsi-Ui Ik, Wen-Chin Peng, (2020) “Trajectory-based Badminton Shots Detection, ” ICPAI, pp.64-71.

Yui Yoshikawa, Hidehiko Shishido, Masashi Suita, Yoshinari Kameda, Itaru Kitahara, (2021) “Shot detection using skeleton position in badminton videos, ” SPIE Proceedings, Vol.11766, IWAIT.

Shayne Vial, Jodie Cochrane, Anthony J. Blazevich and James L. Croft, (2019) “Using the trajectory of the shuttlecock as a measure of performance accuracy in the badminton short serve,” International Journal of Sports Science & Coaching, Vol.14, pp.91-96.

Rusdiana A. 3D Kinematics Analysis of Overhead Backhand and Forehand Smash Techniques in Badminton. Ann Appl Sport Sci. 2021; 9 (3).

Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, (2019) “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 pages.

K. Sun, B. Xiao, D. Liu, J. Wang, (2019) "Deep High-Resolution Representation Learning for Human Pose Estimation, " CVPR, pp.5693-5703.

Dejan Stepec, Danijel Skocaj, (2022) “Video-Based Ski Jump Style Scoring from Pose Trajectory,” IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, pp.682-690.

Naoki Nonaka, Ryo Fujiwara, Monami Nishio, Hidetaka Murakami, Takuya Tajima, Mutsuo Yamada, Akira Maeda, Jun Seita, (2022) “End-to-End High-Risk Tackle Detection System for Rugby,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp.3549-3558.

Kaustubh Milind Kulkarni, Sucheth Shenoy, (2021) “Table Tennis Stroke Recognition Using Two-Dimensional Human Pose Estimation,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp.4571-4579.

Wei-Ta Chu, Samuel Situmeang, (2017) “Badminton Video Analysis based on Spatiotemporal and Stroke Features,” ICMR, pp.448-451.

F. Josue, M. F. Abdullah, I. Zulkapri et al., (2020) “Movement pattern in term of court coverage among top international male and female badminton players during BWF World Championships 2013,” Jurnal Sains Sukan & Pendidikan Jasmani, vol. 9, no. 1, pp. 9–14.

Careelmont S, (2013) “Badminton shot classification in compressed video with baseline angled camera,” Master Thesis, University of Ghent.