Heatmap Visualization and Badminton Player Detection using Convolutional Neural Network

Muhammad Abdul Haq
Faculty of Systems Design
Tokyo Metropolitan University
Tokyo, Japan
muhabdulhaq@gmail.com

Shuhei Tarashima
Innovation Center
NTT Communications Corporation
Tokyo, Japan
tarashima@acm.org

Norio Tagawa
Faculty of Systems Design
Tokyo Metropolitan University
Tokyo, Japan
tagawa@tmu.ac.jp

2022 International Electronics Symposium (IES) | 978-1-6654-8971-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IES55876.2022.9888717

***Abstract*— Badminton coaches and analysts are more interested in how well athletes do in games by watching video matches. However, they still keep watching the whole video by hand, which is inconvenient and might cause them missing important information in the video. Most studies for sports video analysis have been done on soccer and volleyball, but badminton has not been fully focused on. Based on this observation, in this work we aim to build an automated system that can track the position of a player from an input badminton broadcast video, and visualize its position statistics on a heatmap. Convolutional neural network is used to track players and their position is projected on 2D court map using homography. In this paper we validate our approach using videos collected from the Badminton World Federation (BWF) channel on YouTube.**

Keywords*— *multiple object tracking, convolutional neural network, homography

I. INTRODUCTION

Sports videos are rich resources for coaches and analysts to reorganize players' action and decision and improve player performance by examining previous games. To this date, most of them still employ traditional methods of manual video notation, which require a long time to check and result in overlooking important information in the video. For professional coaches and analysts, there are several choices for sports video analysis. However, they are usually costly and need licensing, which are not easily available for amateurs. Moreover, these systems are focused on well-commercialized sports like soccer and volleyball. For other sports like badminton, technical staffs still struggle to analyze large amount of videos manually. Typically, the target audience for sports analytics consists of coaches and community organizations[1].

Based on this motivation, in this work we propose an automatic analysis system tailored for broadcast badminton videos. Specifically, from an input badminton video we track players by using a state-of-the-art convolutional neural network (i.e., YOLOv3) and multiple object tracking algorithm (i.e., SORT). Also, for coached and analysts to investigate player positioning, we propose to generate a position heatmap of players via projecting players bounding boxes into 2D court map by homography. To our knowledge, our work is the first trial work on recognizing players in broadcast footage of badminton matches and visualize their positions by heatmap style. We believe our algorithm provide information about the player's movement and assisting coaches in evaluating current performance, and immediately boosting the player's skills. In this paper we validate our approach using videos collected from the Badminton World Federation (BWF) channel on YouTube. From our evaluation, we validate the effectiveness of our heatmap-based visualization to analyze player performance.

Here are the sections of the study. The existing literature on badminton and player identification, tracking, and analysis is reviewed in Section 2. The third section discusses the materials that were used to attain the result. The fourth section we show the result of the research. Finally in the section 5, the potential for future advances is highlighted

Content analysis is a necessary step to automatically extract information from unstructured video data. Various video analysis techniques such as movement classification, player identification and activity categorization can be an answer to fans' and experts' requests, because they can be transformed into match statistics [2].

For badminton, there are three main tasks of interest in video analysis: shot classification, highlight extraction, and tracking of the ball and the players involved.

Fig. 1. Badminton is a court or lawn game played with shuttlecocks and rackets by two or four players. It recognizes single and double badminton that may be exclusively male, female, or mixed in professional sports.

II. RELATED WORKS

The identification and tracking of players in sports video has been proposed in several previous works, while others have discovered that content characterization in sports programs can be accomplished using a variety of techniques based on low-level features like audio, video, and text captions that are extracted from the video. For tracking the players position in badminton, statistical theories are among the most prevalent methods used to extract data from signals[3]. When compared to traditional machine learning methods, deep learning is believed to be the most promising[4]. Compared to the previous approach which use shallow knowledge, deep learning requires considerable amount of data especially in the case of supervised learning. To this date, variety of deep learning approaches have been proposed by many researchers and the model sizes are getting larger, which consists of a number of hidden layers.

III. Materials and Methods

In this section, we will explain the method that we use for this research. From the video file as in put, object detection, multiple object tracking, projection by homography, and we perform heatmap generation to generate heatmap visualization. We will also explain the basics of Convolutional Neural Networks (abbreviated as CNN), which are one of the most popular approaches of deep learning for video analysis.

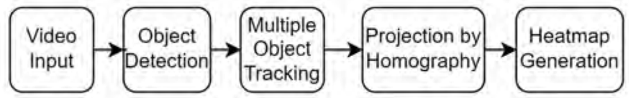

Fig. 2. Program stages to produce heatmap visualization from inputting video file until heatmap generation process.

Further, we introduce the concepts of multiple object tracking, and the many performance measurements utilized to measure player performances.

A. Convolutional Neural Network

It is possible to use a pooling and non-linearity layer in addition to the input, output, and convolutional layers in the CNN-based classification architecture [5]. The classification output is provided by a regression layer that includes one or more fully-connected layers. Both the output and input layers of the design have the same number of neurons, and the input picture has the same size.

In order to extract features from the layer prior to initializing convolutional layers, the weights are utilized. Gaussian weight matrix (F x F) and picture segments of the same dimensions are used to determine the convolution value(x).. The final feature detection filter for all layers is generated after the weight matrices for each layer have been changed one by one during the training phase. A significant amount of data makes it impossible to get all of the previous layer's inputs to the current layer's neuron (FxF). The filter's depth is normally greater, while its width and height are often lower. Activations are created via convolution of the filters with the preceding layer.

B. Multiple Object Tracking

Object detection and tracking are two key techniques to build our system. The process of finding the position and categorization of an item in each picture is known as object detection. In order to track an item in real time, it must first be detected in the video and its exact position must be determined in each frame. In addition to the location of an object's bounding box, these methods also reveal what kind of object it is.

There are a number of CNN models for object detection like Fast R-CNN, Faster R-CNN, Histogram of Oriented Gradients (HOG), etc.. In this work we employ YOLOv3, which delivers accurate detection results while keeping high efficiency[6].

C. Homography

Homography is used to scale the players and create data such as distance and heatmaps[7]. As the points of view change, the analysis may become erroneous. Thus, we must correctly scale down the situation. Homography aids us in establishing a standard or ideal perspective of the field and players in this regard. The goal is to provide coaches and players with a visual analysis that they can link to the game field or individual players.

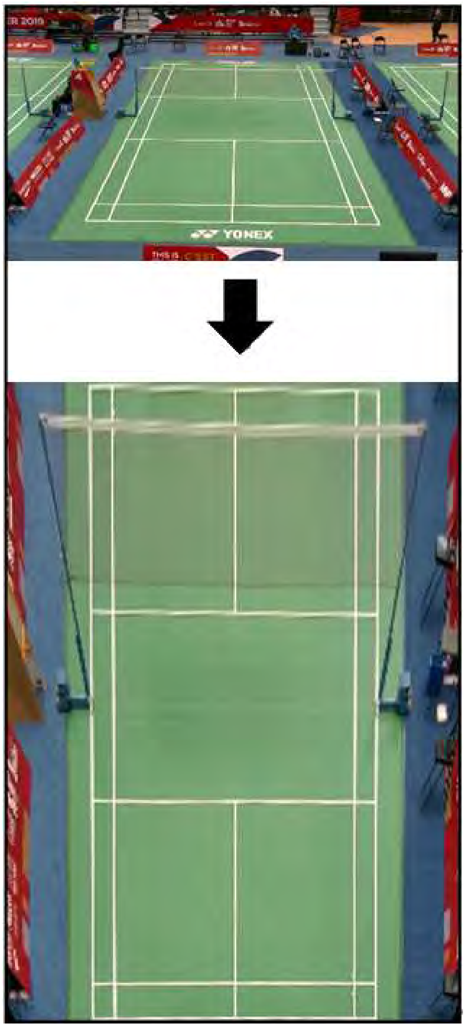

Fig. 3. With homography, we can analyze matches more easily.

IV. Results and Discussion

This section explains the frameworks and deep learning models we employed in this article in great depth. The work at hand is broken down into two parts: player detection and tracking and ball detection and tracking.

A. Detection using YOLOv3

An end-to-end neural network, the YOLO (You Only Look Once) model, finds things in an image using bounding box coordinates and their class. An algorithm known as logistic regression is used to estimate the objectness of each bounding box [8]. There are a lot of tiny cells in the input frame with this procedure. In addition to the box's height, width, x, and y coordinates, this cell also contains an indicator of the object's confidence score and the bounding box center's confidence score. Other levels utilize ReLU activation, however the final layer of YOLO adopts a linear model.

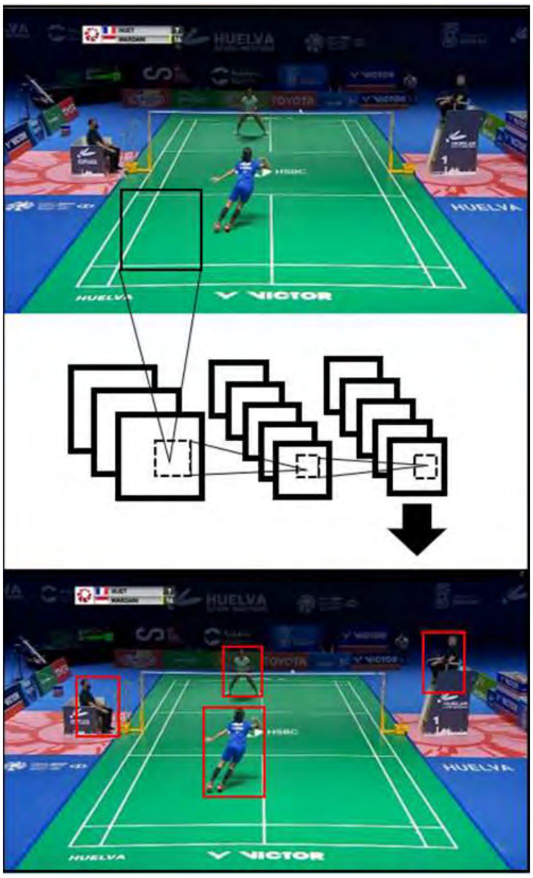

Fig. 4. A Convolutional Neural Network's (CNN) fundamental design. The graphic depicts the different layers present, such as conv, pooling, and ultimately connected layers.

The YOLOv3 architecture is used to recognize players in this paper, and it has been pre-trained using the COCO dataset. This model was created by improving the YOLOv2 model further. There are bounding boxes called anchor boxes in the YOLOv2 model as well as other object recognition models like Faster-CNNs, and these boxes are changed during training. There is no direct correlation between the YOLOv2 model and the item's size and placement in the frame (in terms of the bounding box). The predefined bounding boxes may be moved and resized using an offset instead (anchor boxes). Using a logistic regression algorithm, the outcomes of this reshaping are produced while keeping an eye on a grid cell.

For each of the bounding boxes, the YOLOv3 network predicts four coordinates, and the predictions are as follows:

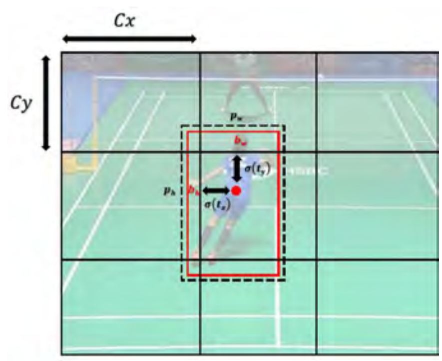

Fig. 5. The width and height of the identified item are contained in the form of the bounding box.

If you want to know how far from the top-left corner a cell is in the image, then (cx, cy) is the cell offset. Loss, on the other hand, is computed as the sum of all squared errors during training, which indicates the difference between ground truth and model predictions.

Each bounding box is predicted to have a "objectness score," with 1 being the most overlap with a ground truth object that can be predicted using Logistic Regression (YOLO v3). If the overlap isn't perfect, but it's higher than a certain threshold, the forecast is ignored. A binary cross-entropy loss is used in the training phase, and a multi-label classification is implemented when more than one class appears inside the bounding box. From the YOLOv3 network, features may be extracted using a feature pyramid network, which employs three separate scales to predict the number of boxes in a picture. Finally, a final convolutional layer predicts a 3D tensor encoding bounding box with a bounding box, class predictions, and objectness.

When the two preceding layers are upsampled by two, the feature map is formed. Fine-grained and relevant semantic data is subsequently generated by concatenating these features with a feature map from inside the network. The model repeats this method in order to predict the final scale bounding box. Bounding box priors are computed using the method called K-means clustering. Instead of using Euclidean distance, a more typical k-mean strategy makes use of a smaller bounding box, which leads to increased bounding box error.

Fig. 6. Two players were detected in a single match, Tai Tzu-Ying with index ID 0 and Akane Yamaguchi with index ID 1.

B. *Tracking*

1) Hungarian Algorithm: When tracking objects, the Hungarian method just uses the YOLO model's bounding box coordinates, disregarding any other video input cues that may be present. Euclidean distance between observed and expected bounding boxes is calculated by comparing the observed and anticipated centroids, the cost function estimates how much an item will cost. For the same item, it calculates the change in size between the current bounding box and the one allocated earlier.

2) Simple Online and Real-time Tracking (SORT): The tracking of a small number of characteristics retrieved from the frame, such as corner points, is commonly done.

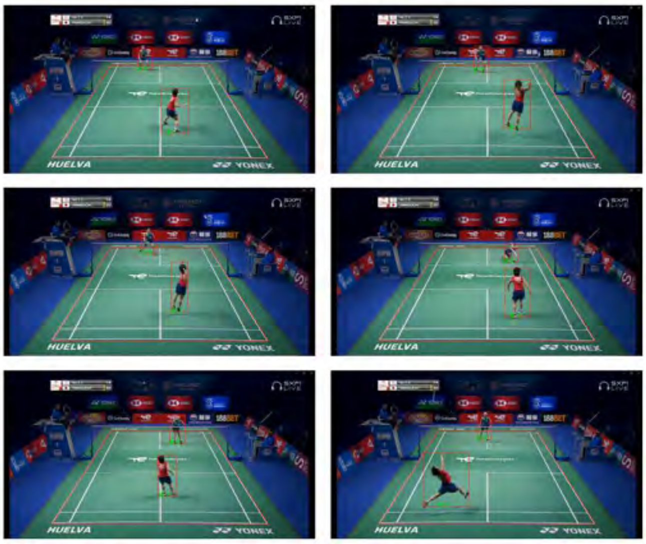

Fig. 7. The result of the tracking system is shown in some frames. The system accurately detected players.

C. Court Homography

Lines on the ground plane determine the playfield's geometry, which is reflected in the court model. A set of vertical lines and a set of horizontal lines are used to organize them. The groups are arranged from top to bottom and left to right. After sorting, the correspondences between the court model lines and the detected lines are identified by iterating all the quadrangles that follow the indices order. These quadrangles can then be used to determine the four intersections.

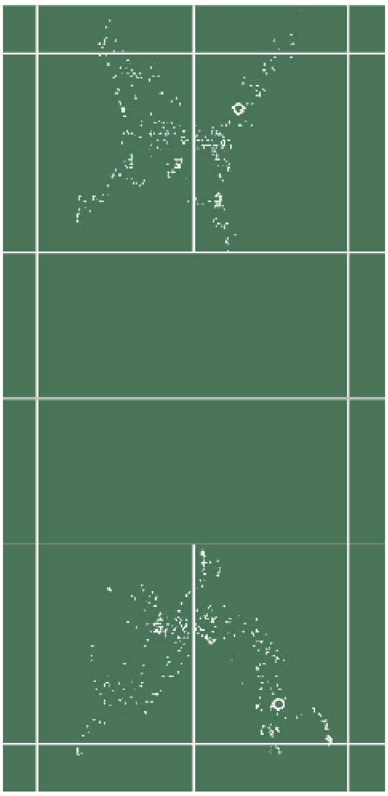

Fig. 8. Homography result after scaling and visualize in badminton court model.

The homography is computed, the court lines are projected to the image plane domain based on this homography, and the matching between the extracted lines and these projected lines is calculated. The calibration points are chosen from the four corners of the assignment with the highest matching score.

D. Heatmap Visualization

After homography, we obtain the frequency matrices of each participant. We make a customized heatmap by choosing the color spectrum that must be shifted from low to high. The downward transition color indicates that the player has only visited that region of the court a few times. For the high transition color, it's the same as it is for the low transition color.

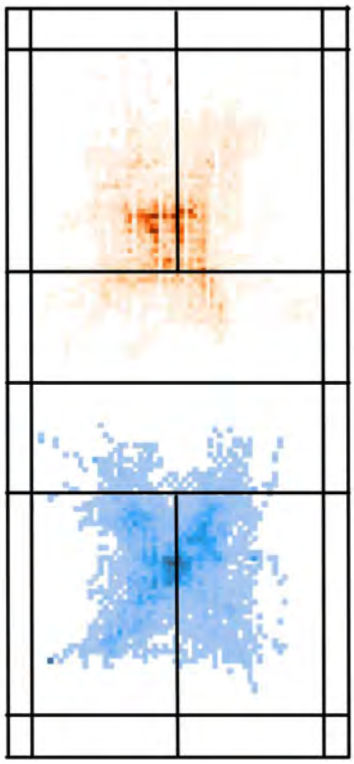

Fig. 9. The final result of heatmap visualization between two players in a single match.

The heatmap shown in figure 9 is the match between Loh Kean Yew from Singapore (orange) vs. Anders Antonsen (blue) from Denmark in BWF World Badminton Championship 2021, the winner is Anders Antonsen. From figure 9 we can see the visualization result. With this heatmap, coaches and players are expected to be able to analyze what deficiencies can be improved, especially in the positioning area. On the other hand, with the results of this heatmap, coaches and players can analyze the movement habits of opponents. The match video was taken from the BWF YouTube channel. The heatmap result is the output of one set of a single match

V. Conclusion

The visualizations and fast results can help players improve their strategies and their performance. It also aids in identifying holes in their opponent's game. There includes a detailed study, visualizations, and explanation of the issues that have been encountered. The accuracy of player tracking has improved. The majority of the current model's implementations are automated, allowing them to be completed with higher speed and precisions.

REFERENCES

[1] W.-Y. Wang, H.-H. Shuai, K.-S. Chang, and W.-C. Peng, "ShuttleNet: Position-aware Fusion of Rally Progress and Player Styles for Stroke Forecasting in Badminton," Dec. 2021, doi: 10.48550/arxiv.2112.01044.

[2] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," *Appl. Sci.*, vol. 12, no. 9, p. 4429, Mar. 2022, doi: 10.3390/app12094429.

[3] Y. Lyu and S. Zhang, "Badminton Path Tracking Algorithm Based on Computer Vision and Ball Speed Analysis," *J. Sensors*, vol. 2021, 2021, doi: 10.1155/2021/3803387.

[4] N. A. Rahmad, N. A. J. Sufri, N. H. Muzamil, and M. A. As'ari, "Badminton player detection using faster region convolutional neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 3, pp. 1330–1335, Jun. 2019, doi: 10.11591/IJEECS.V14.I3.PP1330-1335.

[5] B. Thulasya Naik, M. Farukh Hashmi, C. Author, and M. Farukh Hashmi mdfarukh, "Ball and Player Detection & Tracking in Soccer Videos Using Improved YOLOV3 Model," 2021, doi: 10.21203/rs.3.rs-438886/v1.

[6] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, doi: 10.48550/arxiv.1804.02767.

[7] "Automated Visual Tracking and Live Data Analysis in Badminton | International Journal of Advanced Science and Technology." http://sersc.org/journals/index.php/IJAST/article/view/17819 (accessed Jun. 07, 2022).

[8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, Accessed: Feb. 22, 2022. [Online]. Available: https://arxiv.org/abs/1804.02767v1.