Heliyon 8 (2022) e10089

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Vision-based movement recognition reveals badminton player footwork using deep learning and binocular positioning

Jiabei Luo [a,1], Yujie Hu [b,1], Keith Davids [c], Di Zhang [a], Cade Gouin [b], Xiang Li [a,e,f,*], Xianrui Xu [d,**]

[a] School of Geographic Sciences, Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200241, China
[b] Department of Geography, University of Florida, Gainesville, FL 32611, USA
[c] Sport & Human Performance Research Group, Sheffield Hallam University, Sheffield, UK
[d] School of Economics and Management, Shanghai University of Sport, Shanghai 200438, China
[e] Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, East China Normal University, Shanghai 200241, China
[f] Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, East China Normal University, Shanghai 200241, China

ARTICLE INFO

Keywords:
Coordination
Badminton player trajectories
Computer vision
Binocular positioning
Deep learning

ABSTRACT

Coordinating dynamic interceptive actions in sports like badminton requires skilled performance in getting the racket into the right place at the right time. For this reason, the strategic movement and placement of one's feet, or *footwork*, is an important part of competitive performance. Developing an automated, efficient, and economical method to record individual movement characteristics of players is critical and can benefit athletes and motor control specialists. Here, we propose new methods for recording data on the footwork of individual badminton players, in which deep learning is used to obtain image coordinates (2D) of their shoes and binocular positioning to reconstruct the 3D coordinates of the shoes. Results show that the final positioning accuracy is 74.7%. Using the proposed methods, we revealed inter-individual adaptations in the footwork of several participants during competitive performance. The data provided insights on how individual participants coordinated footwork to intercept the projectile, by varying the distance traveled on court and jump height. Compared with visual observations by biomechanists and motor control specialists, the proposed methods can obtain quantitative data, provide analysis and evaluation of each participant's performance, revealing personal characteristics that could be targeted to shape the individualized training programs of players to refine their badminton footwork.

1. Introduction

Skilled performance in sports like badminton requires coordination of dynamic interceptive actions to get the racket in the right place at the right time to intercept the shuttlecock (projectile) and defend or attack court space [1]. Competitive sport performance at recreational and elite levels requires players to switch quickly between dynamic states of movement organization in order to cover space on the court and intercept the shuttlecock. To coordinate dynamic interceptive actions, a player's footwork is an important part of performance preparation and athlete development in badminton [2]. As in many other sports, current understanding of badminton footwork is largely based on models gained from the long-term observations of athletes in competition and accumulated experiential knowledge of badminton coaches. However, traditional, non-quantitative methods may suffer from subjective bias. Long-term observations of athlete movement trajectories in competition are time-consuming and laborious, potentially resulting in errors, omissions, and misunderstandings. To improve data collection methods, some researchers have installed force sensors in participant shoes to record data during the lunge by badminton players, analyzing mechanical data on their footwork. Through these mechanical analyses, investigations have revealed the impact and potential damage to the lower limbs of players during lunging movements. For example, Valldecabres et al. [3] found that plantar pressure transferred to the medial side of the forelimb and midfoot when players lunged with non-dominant limbs when fatigued in a study of 13 class A badminton league players. Lam et al. [4] found that repetitive movements (RM) when lunging produced a smaller load rate on the knee than a single movement (SM) lunge. Hong et al. [5] found

* Corresponding author.
** Corresponding author.
E-mail addresses: xli@geo.ecnu.edu.cn (X. Li), xuxianrui@sus.edu.cn (X. Xu).
[1] Jiabei Luo and Yujie Hu contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2022.e10089
Received 18 October 2021; Received in revised form 19 January 2022; Accepted 21 July 2022

that a lunge to the left and front showed a higher vertical impact than other lunge directions. Kuntze et al. [6] not only found that the lunge accounted for 15% of all movements in badminton singles, but also analyzed three badminton-specific lunge tasks (kick, step-in, and hop lunge) using video analysis. Remote sensors can not only be used to detect footwork, but also be placed in the badminton racket to analyze players' stroke play. Wang et al. [7] proposed a specific adaptive feature extraction block to improve the performance of a convolutional neural network (CNN) in badminton motion recognition. Their accuracy in motion recognition was as high as 98.65% for classification of ten strokes based on sensor data. Ramasinghe et al. [8] used manual annotation to extract athlete regions from the video and calculate their HOG (histograms of oriented gradients) features [9] for machine learning and stroke recognition. Their methods were able to classify performance of four different strokes with 98.43% accuracy. Compared with long-term observations of athlete movement trajectories, these methods provide more detailed data reference. Nevertheless, previous studies have not attempted to measure spatial location attributes of athletes, which is of critical importance for sports performance analyses. Development of automated, technological methods for performance analysis to accurately and economically record athlete on-court trajectories during performance is critical for designing skill acquisition and strength and conditioning programs in sports like badminton [10, 11].

To enhance skill, expertise, and development of badminton players it is important to ensure that technological developments are based on strong theoretical principles of motor learning [12]. In this study we used an ecological framework to investigate how badminton players solve specific movement problems by coordinating actions during performance and exploiting movement pattern variability [10, 13, 14]. An ecological rationale emphasizes a *process-oriented* approach to performance analytics and technological innovations in sports practice, rather than being driven by data on *performance outcomes* alone [15]. A process-oriented approach prioritizes an analytic focus on skill performance and movement organization in practice and competition by individual athletes, rather than frequency analyses measuring performance outcomes only. Such an approach preferences an individualized analysis, as recently adopted by Giménez-Egido and colleagues [16] in observations of performance of children aged 10 yrs during junior tennis competitions. Their evidence revealed the importance of ensuring that learners experience significant amounts of variability in practice designs to enhance their capacity to adapt their tennis skills to variations in competitive performance.

Here, we developed and implemented an individualised, technological approach to understanding coordination of footwork in another racket sport, badminton, using motion analysis.

Device-based motion analysis is mainly based on wearable sensors placed in shoes, exemplified by attempts to record angular acceleration of a runner's calf during a race [17] and by adding a pressure gauge to shoes to calculate heel and foot pressure [18]. However, these device-based motion studies only addressed what an individual participant is doing, but not their transitions in time and space during performance.

Vision-based motion analysis [19] has also been widely used to study performance in many sports. For example, Nepal et al. [20] recognized goal events from a video clip of a basketball game by using feature extraction. Urtasun et al. [21] used monocular cameras to track the golf swing mode used by participants. He et al. [22] extracted the three-dimensional volleyball trajectory from a game video using color tracking and 3D space matching. Host et al. [23] were able to distinguish multiple players by their performance in handball games by using a multi-target tracking algorithm. Guo et al. [24] designed a two-stage cascade CNN model to judge the membership relationship of hockey players by identifying the color of players' jerseys. Ren et al. [25] invented an innovative algorithm to estimate the three-dimensional trajectory of football using multiple fixed cameras. These studies employed the traditional method of motion analysis—observing the game video with naked eyes—which is inefficient and prone to bias and error. As computer vision technology is progressively improved, it has started to replace human vision in various applications, and the number of cameras involved in positioning has gradually increased. However, current research on vision-based motion analysis typically focuses on overall positioning of performers during competition. Computer vision systems seem ideally placed for determining transitions in (re)organization of important parts of an athlete's body, such as the feet, during competitive performance.

Currently, most of the research examines a two-dimensional (2D) plane and lacks three-dimensional (3D) information, which is an important element in the analysis of dynamic interceptive actions like badminton. Rahmad et al. [26] introduced a badminton player recognition method based on Fast-RCNN (Fast Region-based Convolutional Neural Network), whose recognition object was the whole badminton player, with no calculations of the actual positioning of the badminton player given. Shan et al. [27] used a wireless inertial sensor system to study the dynamic data of upper limb movement including wrist, elbow, and shoulder in the process of hitting the shuttlecock. His research did pinpoint specific body parts of an athlete, but the sensors were unstable compared to vision-based analysis. Currently, there has been no research proposing a vision-based method to extract the 3D trajectory of body parts of badminton players during competitive performance.

In order to calculate the 3D trajectory coordinates of badminton players during performance, binocular vision is needed, which further requires the image coordinates of the same object in two or more cameras [28]. There exists much research analyzing badminton shuttlecock trajectories on court, using 3D positioning methods. For example, Shishido et al. [29] proposed a method using the coordinates of badminton shuttlecock in multiple view planes to calculate the 3D position of badminton shuttlecock using multiple-view videos. Lee [30] used similar methods for 3D positioning of a badminton shuttlecock to automatically analyze players' tactics and predict game outcomes. Based on their work, the University of Science and Technology Beijing robot team designed a robot [31] which can automatically hit a badminton shuttlecock based on 3D positioning data obtained through the camera installed on the robot. Furthermore, Chen et al. [31] proposed clock-synchronization, combined with motion compensation methods, to improve localization error. Although these methods only typically focused on the 3D trajectory and location of a badminton shuttlecock, the idea of using multiple cameras for obtaining 3D information of a spatial entity remains an inspiring possibility for future research. The remaining problem is how to efficiently extract the coordinates of an object from the frame of video clips shot from different angles.

Perhaps the first study in the field that improved the detection accuracy of objects, such as cars, animals, and person, was by Girshick et al. [32], which used the Region-based Convolutional Neural Network (R–CNN). However, the process of selecting ~2,000 candidate frames did not employ a shared convolutional network for calculation, resulting in a slow detection speed. Later, He et al. [33] proposed the Spatial Pyramid Pooling Net (SPP-Net) algorithm to increase the rate of generating candidate frames by adding Spatial Pyramid Pooling (SPP) to both the convolutional layer and the fully connected layer. Fast R–CNN [34] has since been developed based on SPP-Net, replacing the SPP layer with the ROI (Region Of Interest) Pooling layer and sharing the convolutional layer in the entire network. Although this approach allows the fully connected layer network to perform regression and classification tasks at the same time, thus greatly reducing training and detection time, Fast R–CNN still needs to generate a large number of candidate regions, which demands much computation. Ren et al. [35] proposed the Faster R–CNN algorithm to solve this problem. With the development of deep learning, it allows for a framework where the pixel position of an object in the images captured by two cameras can be obtained and matched for binocular positioning. Monezi et al. [36], for example, reconstructed 3D positions of basketball players (their heads, to be more precise) using deep learning and binocular positioning. Their focus was simply to extract discrete 3D positions of basketball player heads and they did not

further their analyses by extracting players' trajectory and analyzing their performance.

Thus, the overall goal of this paper is to propose automated, efficient, and economical methods to extract the 3D trajectory of individual badminton players, during competitive performance, using deep learning and binocular positioning methods. Here we sought to demonstrate the specificities of movement trajectories of individual players, varying in key characteristics such as sex differences, age, and skill level, to investigate the utility of the proposed method in sport motion analysis. For this purpose, we examined how individual badminton players varied their footwork, based on personal characteristics, when seeking to get the racket into the right place, at the right time, to intercept a projectile during competitive performance.

The paper is organized as follows. Section 1 describes the background, reviewing relevant literature in the scope of the paper. Section 2 details the data sources and models used in this study to extract data on badminton players' movement trajectories. Section 3 presents the model performance by using a series of evaluation metrics, while section 4 applies the methods to study inter-individual adaptations in the footwork of badminton players during competitive performance. Finally, section 5 concludes the paper.

2. Dataset and methods

We mounted two cameras in a badminton court (refer to section 2.2 for their locations in the court) to capture the video footage of the badminton player using the proposed methods. The first camera (Camera A) is a Canon EOS 6D Mark II and the second (camera B) is a Canon EOS 77D. Both cameras have a frame rate of 25 frames per second. World coordinates of cameras A and B (their optical axis centers) are (3, −2.9, 1.4) and (12.3, 9.34, 1.5), respectively. In our study, we sampled two video clips using each camera—one lasting 30 s (750 images) the other 71 s (1775 images). Specifically, of the first dataset (1,500 images in total which is illustrated as *Image Sequence of Camera A* and *Image Sequence of Camera B* in Figure 1), 90% of the visual information (1,350 images) was used for model training. The remaining 10% (150 images) was used for model validation. From the second dataset, 352 images (about one in every five frames) were sampled for shoe localization and trajectory extraction. Appendix C (a) and (b) shows the 90th frame in the second dataset associated with cameras A and B, respectively. The study was approved by the University Committee on Human Research Protection of East China Normal University. Written informed consent was obtained from all individual participants included in this study.

The process of the proposed methodology is composed of three major steps. In step 1, we applied a deep learning model to identify coordinates (2D) of a player's shoes (i.e., shoe localization) in images captured by cameras. In step 2, we proposed methods to convert image coordinates of shoes to world coordinates (3D), by: (1) defining the world coordinate system based on which locations of any given point on the badminton court can be defined, (2) deriving *x*-*y* world coordinates of shoes based on corresponding image coordinates identified in step 1, and (3), estimating *z* world coordinates of shoes from examining images captured by multiple cameras using the binocular positioning method. Lastly, in Step 3, the footwork trajectory of the player was constructed by connecting discrete shoe data coordinates. Based on the extracted footwork trajectory, we then performed individualized analyses to examine the relationships between players' footwork trajectory and their performance and relevant sociodemographic characteristics. Figure 1 illustrates the workflow of the proposed methodology.

2.1. Shoe localization in images using deep learning

We used the convolutional neural network Visual Geometry Group Network 16 (VGG16) [37] to extract image feature maps. The VGG16 is a pre-trained weight network based on the ImageNet [38] dataset. It contains 13 convolutional layers (Conv) and 4 pooling layers. In order to introduce nonlinear relationships into neurons in the network, each convolutional layer has an activation function called Rectified Linear Unit (ReLU) [39] that makes the training network converge quickly, corresponding to the neuron (Figure 2). For each convolutional layer, the

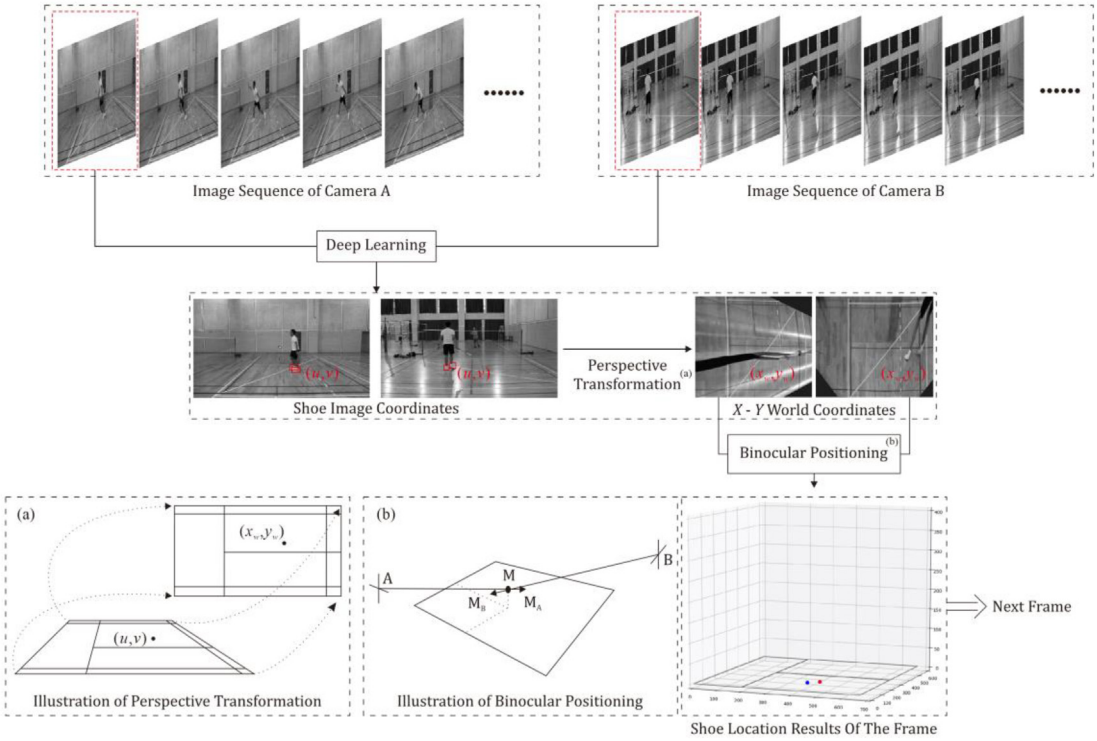


Figure 1. Workflow of the proposed methodology.

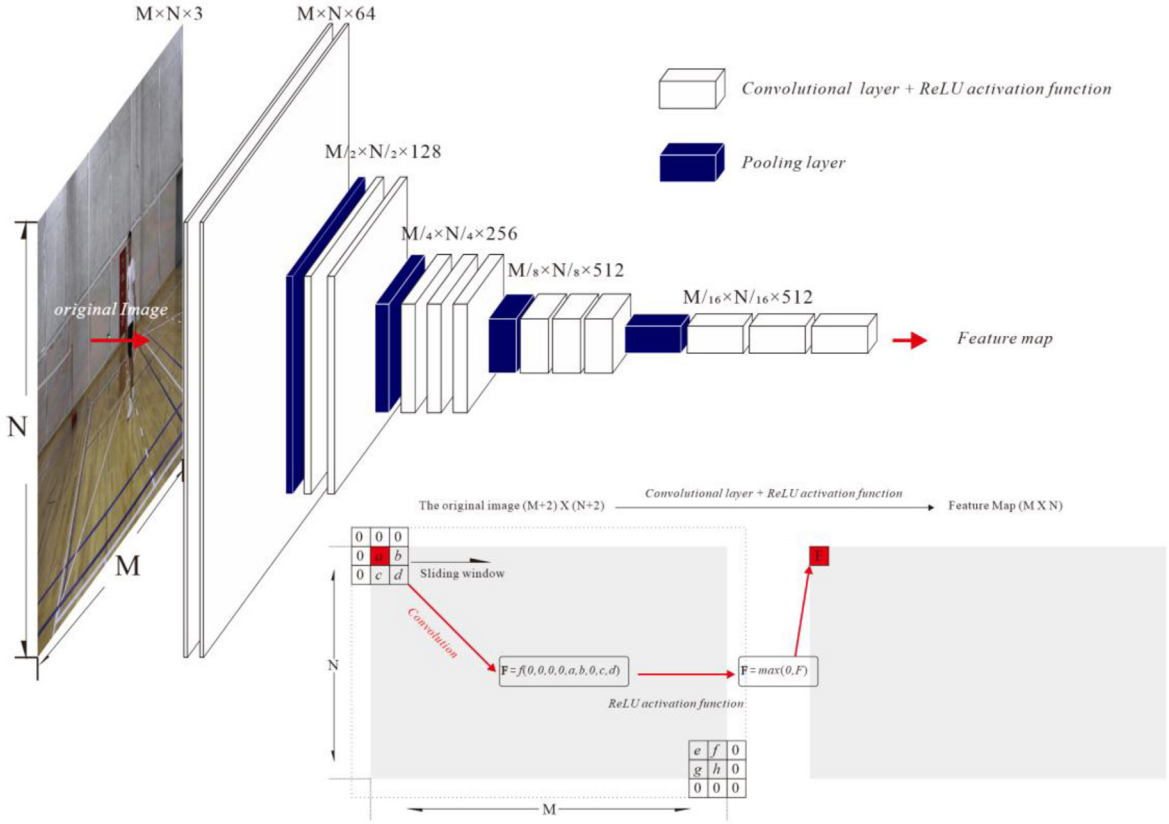


Figure 2. The VGG16 network architecture (adapted from [37]).

size of the convolution kernel is 3 × 3, and the step size is 1. To ensure that there is no omitted information after image convolution, the image boundary is filled with pixel value 0 (pad = 1) in advance. The core size used by the pooling layer is 2 × 2, and the step size is 1. After passing through the pooling layer, the dimension (i.e., the number of convolution kernels) of the feature map is doubled (except for the last pooling layer). The dimension of the feature map grows from 64 to 512. The size of the feature map is reduced to 1/4 of the original image. Since neither the convolutional layer nor the ReLU activation function changes the size of the feature map, for an RGB image with an input size of M × N, the size of the feature map output by the network is (M/16) × (N/16) with the dimension of 512.

The convolution workflow is as follows. First, the image is expanded to the size of (M + 2) × (N + 2) (Figure 2). A 3 × 3 sliding window is then used to convolve the image, and a feature value is calculated using the convolution on 9 pixels in the sliding window. The final feature value calculated by the ReLU activation function is the feature value corresponding to the location of the feature map at the center of the current sliding window. The purpose of the ReLU activation function is to introduce nonlinear relationships into neurons, which can also be called

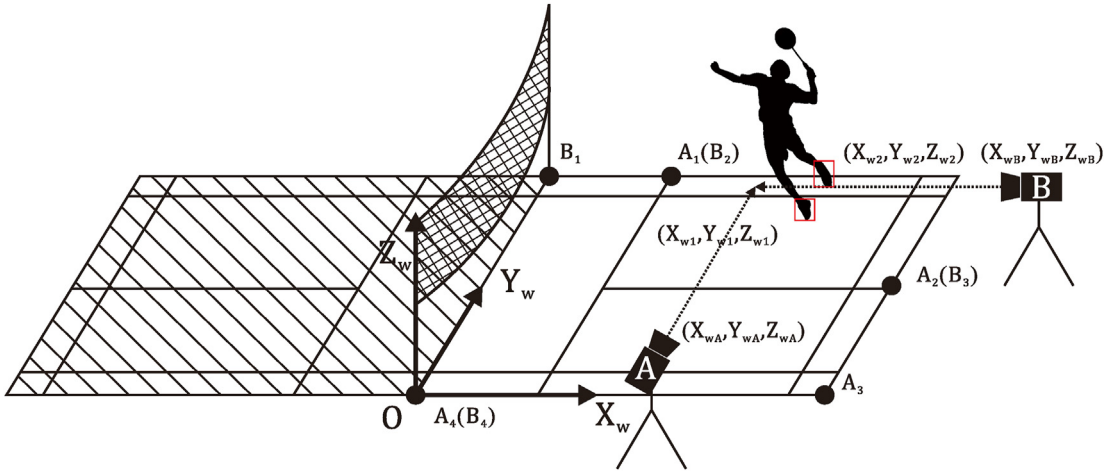


Figure 3. Camera setup and definition of the world coordinate system.

sparse activation, and this is an indispensable part of the convolutional layer operation.

2.2. Converting image coordinates to world coordinates

The shoe location obtained from section 2.1 is essentially image coordinates (2D). Here, several steps are developed to convert a shoe's image coordinates to world coordinates (3D). For simplicity, we demonstrate our methods only by looking at a single player in a half court with two cameras installed—one on the left side of the court and the other behind back court [40] (see Figure 3). For full-court recordings on movements of two players in a specific competitive game, we can simply introduce an additional camera C on the other side of the court and move camera A to the middle of the court. In this way, camera A will obtain coordinates of four shoes associated with two players. Cameras B and C will capture the trajectory of player 2 and 1, respectively, from the back.

Image formation in camera requires transformation of a point's location from its world coordinates to image coordinates. Three steps—rigid body transformation, projection imaging, and pixel coordinate transformation—are involved in this process to convert among three coordinate systems—world coordinate system, camera coordinate system, and image coordinate system. The world coordinate system is a 3D coordinate system based on which locations of points (e.g., player shoes) can be defined. Camera coordinate system is a 3D coordinate system attached to the camera, whereas image coordinate system is a 2D coordinate system specific to the image. Figure 3 illustrates the origin and X_w, Y_w, Z_w axes of the world coordinate system defined in this research. Based on the defined world coordinate system, spatial locations of the optical axis of cameras A and B are given by (x_{wA}, y_{wA}, z_{wA}) and (x_{wB}, y_{wB}, z_{wB}), respectively. The shoe location of a player in the world coordinate system is given by (x_{wi}, y_{wi}, z_{wi}), where i denotes one of the two shoes of a player.

Rigid body transformation relates the world coordinate system and camera coordinate system by a rotation and translation (see Appendix A (a)). Translation transformation is the movement of the origin of the coordinate system denoted by the direction vector T (3×1 matrix). The rotation transformation can be regarded as the transformation of the coordinate axis (x, y, z) denoted by the direction vector R (3×3 matrix). The formula for rigid body transformation is shown in Eq. (1). Its homogeneous expression is shown in Eq. (2).

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2}$$

Once we obtain a point's 3D coordinates in the camera coordinate system by applying a rotation and translation to the point's world coordinates, we can attain the point's image coordinates (location of the point in the image) by projecting the point on the image plane (see Appendix A (b)). Let m (x_{ip}, y_{ip}) denote the point location in the image plane. According to the triangle similarity relationship, $x_{ip}/x_c = y_{ip}/y_c = f/z_c$ can be obtained, where f is the focal length of the camera. Therefore, Eq. (3) can be used to express the process of projection imaging.

$$z_c \begin{bmatrix} x_{ip} \\ y_{ip} \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{3}$$

Since the image coordinates obtained at this stage can only represent the relative position of a point object in the image without the consideration of physical units, we then establish a scaling relationship to further calibrate the image coordinates. Let (x_{ip}, y_{ip}) and (u, v) denote the image coordinates before and after the calibration, respectively. The scaling relationship can be formulated in Eq. (4).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ip} \\ y_{ip} \\ 1 \end{bmatrix} \tag{4}$$

Combining all steps together, we can obtain Eq. (5) to convert between a point's image coordinates and world coordinates.

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{5}$$

where $L = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} = \begin{bmatrix} l_1 & l_2 & l_3 & l_4 \\ l_5 & l_6 & l_7 & l_8 \\ l_9 & l_{10} & l_{11} & l_{12} \end{bmatrix}$

Solving Eq. (5) to get the world coordinates of a shoe requires obtaining values for three variables—u, v, and z_c. The values for u and v can be attained from the deep learning model discussed in section 2.1, but the value for z_c is difficult to obtain. Alternatively, we first only calculate the x-y coordinates (x_w, y_w) of a shoe in the x-y world coordinate system from the image shot by either camera. The two sets of world coordinates (x_w, y_w) of a shoe associated with the two cameras are then used to estimate z_w using binocular positioning.

To only measure a shoe's coordinates (x_w, y_w), we can substitute $Z = 0$ into Eq. (5), which is then simplified as: $\begin{bmatrix} l_1 & l_2 & l_4 \\ l_5 & l_6 & l_8 \\ l_9 & l_{10} & l_{12} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} = z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$, where $\frac{1}{z_c} = w'$, $\frac{x_w}{z_c} = x'$, $\frac{y_w}{z_c} = y'$. This gives $\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} l_1 & l_2 & l_4 \\ l_5 & l_6 & l_8 \\ l_9 & l_{10} & l_{12} \end{bmatrix}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ which can be rewritten as Eq. (6):

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{6}$$

where $\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$ is the Homography matrix, and it includes 8 unknowns (except for h_{33} being a constant value of 1).

To solve the 8 unknowns, 8 nonlinear correlation equations need to be constructed. Therefore, it is necessary to identify four groups of homonymous points in the image and world coordinate systems as control points and substitute them into Eq. (7).

$$x_w = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}};\ y_w = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}} \tag{7}$$

Finally, 8 nonlinear correlations describing h_{11}、 h_{21}、 h_{12}、 h_{22}、 h_{31}、 h_{32}、 h_{13}、 h_{23} and h_{33} are obtained and the Homography matrix can be solved. Using the derived Homography matrix, we can measure world coordinates (x_w, y_w) for any given image coordinates (u, v) (see Figure 1(a)).

Once coordinates (x_w, y_w) of a given shoe are measured from either camera, binocular positioning can be applied to calculate z_w. Suppose that the (x_w, y_w) coordinates of shoe M based on camera A are M_A (x_{mA},

y_{mA}) and the (x_w, y_w) coordinates based on camera B are $M_B(x_{mB}, y_{mB})$. As shown in Figure 1(b), the intersection point of line AM_A and line BM_B would be the actual spatial location of the shoe.

However, the intersection point does not exist between two non-coplanar lines as in this case. Alternatively, the midpoint of the shortest segment between the two lines is regarded as the intersection point in this research (see Appendix B). We have AM_A whose direction vector is $\vec{u} = M_A - A$ and BM_B whose direction vector is $\vec{v} = M_B - B$. Suppose that the distances between each of the two points and the origin are A_d and B_d, respectively (where $0 \leq A_d, B_d \leq 1$), then:

$$\begin{aligned} \vec{s_d} &= A + A_d \cdot \vec{u} \\ \vec{t_d} &= B + B_d \cdot \vec{v} \end{aligned} \tag{8}$$

So $\vec{w_d} = \vec{s_d} - \vec{t_d} = A - B + A_d \cdot \vec{u} - B_d \cdot \vec{v} = \vec{w} + A_d \cdot \vec{u} - B_d \cdot \vec{v}$. If line AM_A and line BM_B are not parallel or coincident, there are only two points A_0 and B_0 that make the line segment A_0B_0 the nearest two points between AM_A and BM_B. The line segment A_0B_0 is also the only line segment perpendicular to the two lines at the same time. So:

$$\vec{w_d} \cdot \vec{u} = 0;\ \vec{w_d} \cdot \vec{v} = 0;\ \vec{w_d} = \vec{w} + A_d \cdot \vec{u} - B_d \cdot \vec{v} \tag{9}$$

Then a group of binary linear equations is obtained:

$$\vec{u} \cdot \vec{u} A_d - \vec{u} \cdot \vec{v} B_d = -\vec{u} \cdot \vec{w_d};\ \vec{v} \cdot \vec{u} A_d - \vec{v} \cdot \vec{v} B_d = -\vec{v} \cdot \vec{w_d} \tag{10}$$

We substitute the variables in Eq. (8) and Eq. (9) into Eq. (10) and get A_d and B_d which are shown in Eq. (11):

$$A_d = \frac{be - cd}{ac - b^2};\ B_d = \frac{ae - bd}{ac - b^2} \tag{11}$$

where $a = \vec{u} \cdot \vec{u}$, $b = \vec{u} \cdot \vec{v}$, $c = \vec{v} \cdot \vec{v}$, $d = \vec{u} \cdot \vec{w_d}$, $e = \vec{v} \cdot \vec{w_d}$. Finally, we get the coordinates of the intersection $I = (\vec{s_d} + \vec{t_d}) / 2$.

2.3. Integration of discrete shoe locations to form trajectory

Using the methods in 2.1 and 2.2, the world coordinates of each shoe can be obtained by analyzing corresponding images from camera A and camera B at each time frame. We can then connect the world coordinates of each shoe at each time frame to derive the player's complete movement trajectory. But the methods discussed previously do not have a mechanism to match shoes across images from different cameras or different time frames. In other words, it is unknown which of the two identified shoes in image 1 corresponds to which of the two identified shoes in image 2. A nearest neighbor matching method is thus proposed to match shoes across images.

To match shoes across images from different cameras using the nearest neighbor matching method, we used s (x_s, y_s) and t (x_t, y_t) to denote the (x_w, y_w) coordinates of the pair of shoes in camera A and p (x_p, y_p) and q (x_q, y_q) to denote the (x_w, y_w) coordinates of the pair of shoes in camera B. We then calculated the Euclidean distances d_{sp}, d_{sq}, d_{tp}, and d_{tq} between each shoe in the camera A image and each shoe in the camera B image and defined the nearest neighbor distance as $D = \min(d_{sp}, d_{sq}, d_{tp}, d_{tq})$. The shoe pair between the two images with the minimum distance D is regarded as the same shoe, and the remaining two shoes are the other same shoe. Note that the same process is also applied to match shoes between adjacent images in time associated with the same camera.

3. Results

In the experiment, we used the TensorFlow-gpu 1.10.0 as the deep learning framework. All the analyses were performed in a Windows 10 environment of Intel(R) Core(TM) i7-8700K CPU @ 3.70 GHz and 32G RAM and a GPU of NVIDIA GeForce GTX 1080 Ti, 11G memory. As previously mentioned, the first dataset that includes 1500 images was used for model training. The iteration (rounds of training), batch size (number of training pictures per iteration), and total epochs of training were set to 150000, 4, and 25, respectively. The model training was optimized by using the rmsprop (Root Mean Square Propagation) algorithm [41]. In addition, we extracted the feature map of the image through the pre-trained model weights (VGG 16) and only trained the classification and regression layers of Faster R–CNN to speed up the training process.

3.1. Results of shoe localization

For each of the 1,500 images in the training/validation dataset, we used the labelImg package (https://github.com/tzutalin/labelImg) to manually draw the bounding box for marking the shoes in images (termed the target box). These images labeled with bounding boxes were then compared with the shoe locations identified by the deep learning model (termed the GT box) to evaluate model performance. Specifically, we calculated the overlapping rate of the target box and the GT box, which is formulated in Eq. (12):

$$IoU = \frac{S_C}{S_A + S_B - S_C} \tag{12}$$

where IoU means the overlapping rate of the target box and GT box, S_A represents the area of GT box, S_B represents the area of target box, and S_C represents the area of the intersection of the target and the GT box.

When the IoU is >0.5, the shoe is deemed successfully detected. Four types of detection results are defined: (1) true positive (TP), the number of positive samples detected as positive samples; (2) false positive (FP), the number of positive samples detected as negative samples; (3) false negative (FN), the number of negative samples detected as positive samples; and (4) true negative (TN), the number of negative samples detected as negative samples. With these four values as parameters, we can define the average accuracy (AP) and the mean of average accuracy (mAP) of model detection results (see Eqs. (13), (14), and (15)).

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$AP = \frac{\sum_n Precision}{n} \tag{14}$$

$$mAP = \frac{\sum_m AP}{m} \tag{15}$$

During shoe localization, we input the feature map and the target box that was randomly generated into the ROI pooling layer (third layer of Faster R–CNN [35]) and transferred them into the full connection layer. Then, the feature graph of the target box was classified and regressed to obtain the exact position of the final target box.

The average computational time cost of the VGG16 network for single-frame image detection was 0.2185 s and the mAP was 0.982, indicating a remarkably high accuracy and good efficiency. To help justify our choice of using the VGG16 network, we also did the same test using another popular convolutional neural network—the ZF network [42] (Zeiler & Fergus Net, which is an improvement on AlexNet by tweaking the architecture hyperparameters, in particular, by expanding the size of the intermediate convolution layer). In terms of the ZF network, the average time cost for single-frame image detection was 0.1911 s and the mAP was 0.953. Given the higher detection accuracy of the VGG16 network (and relatively closer computational time) compared to the ZF network, it makes more sense to use the VGG16 network as the convolutional neural network in the Faster R–CNN deep learning model. Appendix C (c) and Appendix C (d) show the detection results of the 90th frame of cameras A and B, respectively.

3.2. Results of the transformation from image coordinates to world coordinates

As discussed in Section 2.2, four control points are needed for the coordinate transformation. A rule-of-thumb for the selection of quality

control points is that they should be at the intersection of edges and be dispersed as much as possible. See Figure 3 for more detail about the four selected control points (marked in black).

Ten water bottles were randomly placed around the court to evaluate the accuracy of the binocular positioning method. As discussed in Section 2.2, world coordinates of the ten water bottles (bottle caps, to be precise) after binocular positioning were derived (see Appendix D). The error of binocular positioning is defined as the Euclidean distance between the actual world coordinates of the bottle caps and the estimated world coordinates using binocular positioning. The average error was 0.129 m, which, compared to the size of the badminton court (6.7 m $\times$ 6.1 m), indicated high accuracy of the proposed methods.

Then our methods were applied to the second dataset for shoe localization and trajectory extraction. As shown in Figure 4 (dots of different colors represent different shoes of a player), the shoe identification results can be divided into three types: (1) both the shoe location and category (left or right) are accurately identified; (2) only the shoe location is accurately identified but shoe category is reversed; and (3) neither the location nor category of the shoe is accurately identified.

Results showed that both the shoe location and category are accurately identified in 74.7% (263 of 352 images) of the second dataset. In about 22.4% (79 of 352 images) of the dataset, only the shoe location is accurately identified. Only in about 2.8% (10 of 352) of the dataset was the shoe location not correctly identified. If we do not consider the shoe classification error, the overall identification accuracy of the proposed methods is as high as 97.2%, indicating the efficacy of using deep learning and binocular positioning to identify a player's foot locations and derive their footwork trajectory during competitive performance. The extracted footwork trajectory (based on results where both the shoe location and category are correctly identified) of a player is presented in Figure 5, where the lines represent a pair of shoes with blue dots denoting the left foot and red dots denoting the right foot. Note that the

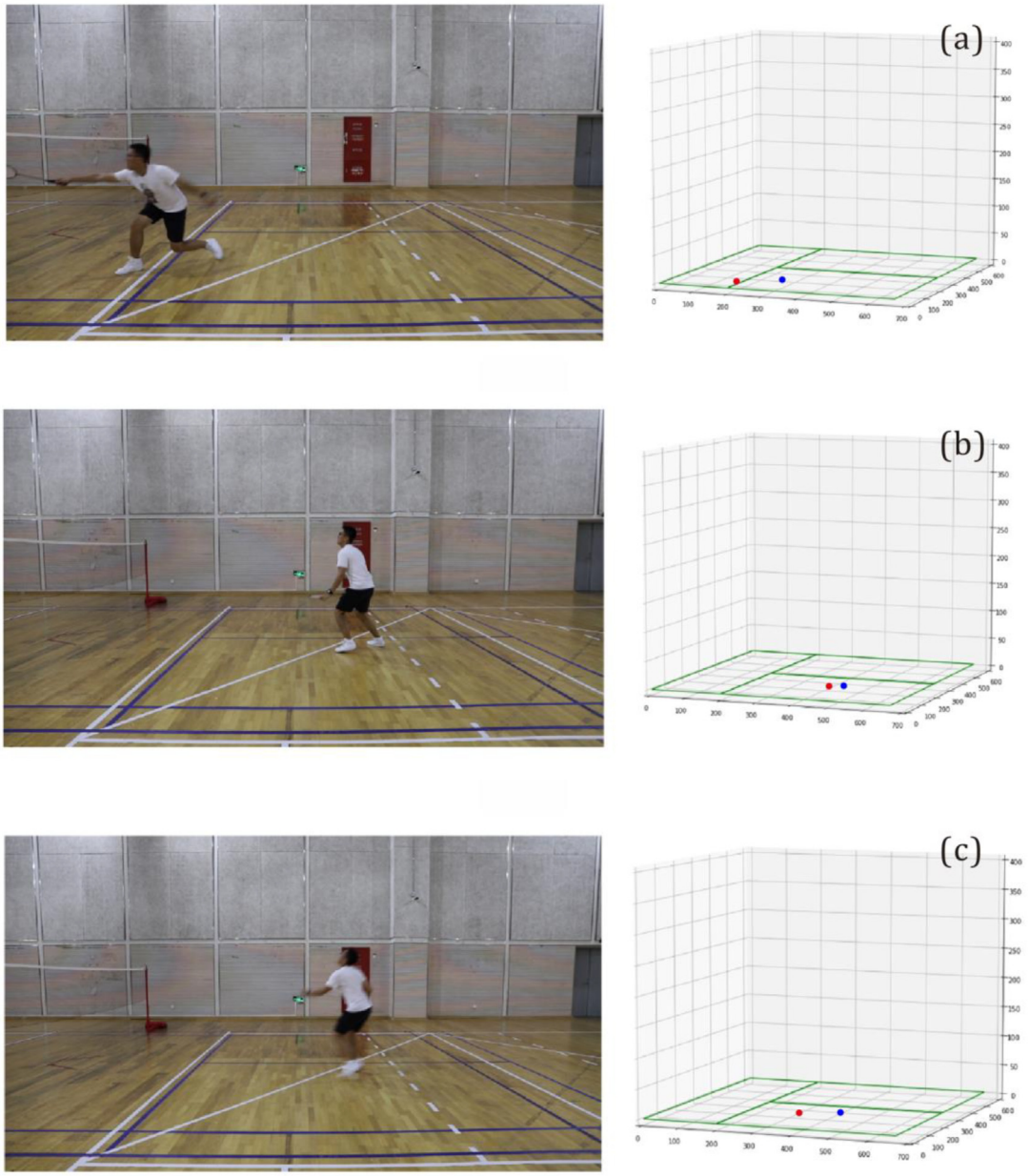


Figure 4. Shoe identification results: (a) both the shoe location and category are accurately identified; (b) only the shoe location is accurately identified; and (c) neither the location nor category of the shoe is accurately identified.

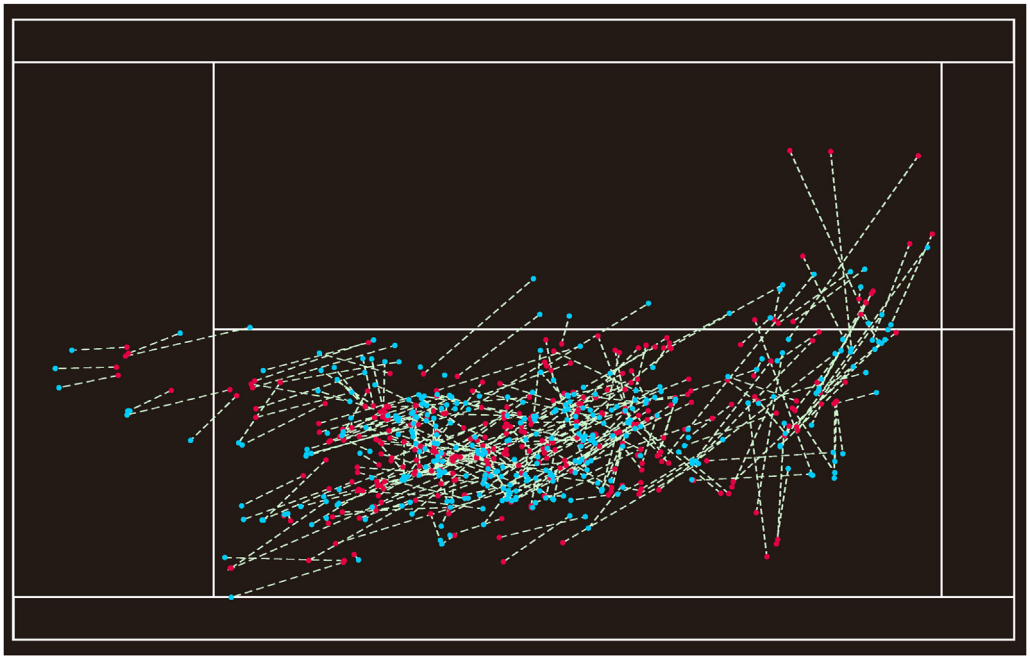


Figure 5. Footwork trajectory of a player.

participant's foot trajectory was mostly on the left side of the court due to their opponent being left-handed.

Our methods show unique advantages when compared with other methods. The methods by Shishido [29] and Lee [30] were able to achieve the 3D positioning accuracy of a badminton shuttlecock at 0.48 m, while the accuracy of our methods for shoe localization is as high as 0.129 m. Huang et al. [43] designed a neural network regressor and combined it with human joint information detection to reconstruct the 3D movement of athletes' limbs, in order to realize the auxiliary training of badminton singles athletes. However, their method did not work in real time, while the data obtained by our method were obtained in real-time analyses.

4. Discussion

In this study, we used an ecological dynamics framework to observe competitive performance in six participants, at the sub-elite level, playing 10 badminton games which lasted for 2 min each (see Appendix E for information about these participants). The data from our computer-vision based performance analytics system showed how sport scientists, biomechanists, and motor control specialists could adopt a more individualized, process-orientation to understanding movement coordination (exemplified by footwork) in sports like in badminton. Several metrics were employed to measure the movement performance of the participants. They included: (1) **total distance moved:** the sum of the Euclidean distances of all adjacent track points of a single shoe; (2) **average bounce height:** the average z value of all track points of a single shoe; (3) **maximum bounce height:** the maximum z value of all track points of a single shoe; (4) **average moving speed:** the moving speed is the Euclidean distance between adjacent track points of a single shoe divided by 0.2 s (25 frames/second in the video), and thus the average value of the moving speed of a single shoe across the whole time period is the average moving speed; and (5), **maximum moving speed:** the maximum value of the moving speed of a single shoe at each moment. Our analyses were individualized, focusing on inter-individual performance variations, predicated on relevant factors including differences in *sex differences*, *age*, *exercise frequency*, *weight*, and *height*, as well as the *score between opponents*, and *number of strokes* and *number of mistakes made*.

4.1. Analysis of individual footwork performance varying by factors

As predicted by an ecological dynamics perspective on performance processes, Figure 6 shows how the computer-vision methodology was able to record key variations in footwork by individuals. These data exemplified how personal and task constraints interacted to adapt values of distance moved and maximum bounce height of both shoes for each participant during all 10 games. Each participant has four bars. The red bar represents the average distance moved in all competitions, while the blue bar represents the average jumping height. The bar on the left represents shoe 1 and the bar on the right represents shoe 2. Overall, among the six participants, distance moved was observed to vary by sex differences, with men's values being significantly higher than women's, except for Participant 3. The age of Participant 3 (46 yrs) varied from other participants who are young adults. Participant 4 is a female, recording a high total distance moved and average bounce height due to her high frequency of exercise. We observed that Participant 1 moved 60 m more than Player 2 in just 2 min, which greatly increases the covering court space and possibility of returning stroke. This individual performance characteristic between participants, discriminated by the methodology, may be in part explained by the data showing that Participant 1 exercises five times a week, while Participant 4 exercises only once a week. The methodology was able to pick up data indicating a positive relationship between exercise frequency and movement distance values achieved during performance between individuals.

The average lines in Figure 6 showed that movement distance and maximum bounce height for males were higher than for females. Participant 1 who won 3 scores per game performed well above the sample average in both movement distance and maximum bounce height. Participant 2 whose movement distance was above average, while maximum bounce height below average, only won 1.5 scores per game, as half as many as Participant 1. The female Participant 4's movement distance and maximum bounce height reached the male average. She won 2.3 points per game, ranking the second out of six participants. The remaining three participants displayed slightly below-average values for movement distance and maximum bounce height, which may be related to their low-ranking competitive performance. In addition, the methodology discriminated that Participant 1 and Participant 2 were able to jump with one foot, compared to other participants

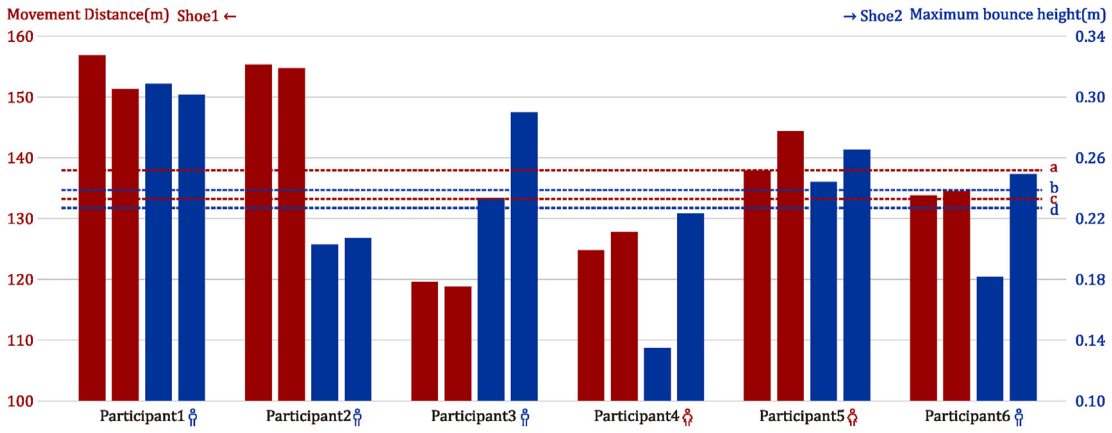


Figure 6. Results of movement distance and maximum bounce height among participants (line a: Average movement distance of males, 137.24 m; line b: Average maximum bounce height of males, 0.237 m; line c: Average movement distance of females, 133.73 m; line d: Average maximum bounce height of females, 0.21 m).

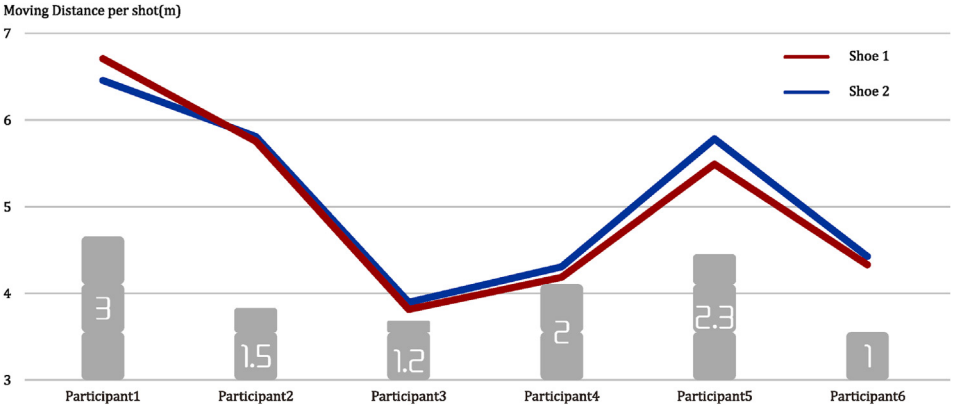


Figure 7. The relationship between distance moved for each shot and winning points.

who could not. This may be the reason why Participant 1 and Participant 2 had a successful 'kill shot' in the game, while the other four players did not. This observation demonstrated the importance of our methods of examining individual shoes.

To highlight the importance of footwork in scoring, we also calculated the average winning score of each individual participant. Figure 7 shows that distance moved for each shot is proportional to the average winning score. The longer the distance moved of each stroke, the easier it

Table 1. Statistics of the games and participants.

	Participants	movement distance of shoe 1 (m)	movement distance of shoe 2 (m)	Maximum bounce height of shoe 1 (m)	Maximum bounce height of shoe 2 (m)	Score	Swing times
Game1	Participant 1	150.1	148.0	0.195	0.184	4	24
	Participant 2	175.8	169.9	0.281	0.257	9	19
Game2	Participant 1	151.3	142.0	0.354	0.393	5	30
	Participant 3	124.8	132.7	0.228	0.417	5	27
Game3	Participant 4	143.4	138.8	0.086	0.185	4	31
	Participant 3	131.8	130.0	0.113	0.245	6	30
Game4	Participant 5	136.9	139.9	0.240	0.252	5	31
	Participant 3	100.4	100.5	0.193	0.178	7	29
Game5	Participant 6	145.7	139.2	0.181	0.248	5	42
	Participant 3	121.9	113.3	0.186	0.343	3	40
Game6	Participant 2	135.0	139.7	0.113	0.141	4	28
	Participant 3	119.3	118.3	0.406	0.209	6	28
Game7	Participant 1	165.8	153.3	0.324	0.283	7	19
	Participant 5	158.6	164.0	0.243	0.270	6	21
Game8	Participant 4	119.6	118.9	0.172	0.221	3	28
	Participant 5	118.2	129.2	0.223	0.243	10	26
Game9	Participant 4	111.3	125.9	0.144	0.244	4	30
	Participant 6	96.2	96.9	0.073	0.116	7	32
Game10	Participant 1	161.4	162.5	0.302	0.291	5	24
	Participant 6	159.8	167.6	0.278	0.356	7	25

is to adjust the appropriate stroke to make a winning hit, which results in a higher winning score. Refer to Table 1 for more detail on competitive performance and individual characteristics of participants. From an ecological perspective, Table 1 captures the process-oriented interactions of personal constraints of each participant and adaptations to task performance as a consequence [44].

For example, the reason why Participant 1 could demonstrate such strong performance in all games was that Participant 1 exercises 1-2 times a week and maintains muscle training. In addition, Participant 1 had attended a badminton course for one semester and studied badminton footwork professionally. Combined with Figures 6 and 7, it could be found that the values for average winning score, movement distance, and maximum bounce height of Participant 1 were the highest. As a female player, Participant 4 had the second highest values for average winning score, movement distance, and maximum bounce height, which were equal to the values of male participants. This finding was perhaps because of her high exercise frequency and being a member of the school volleyball team. Participant 2, whose height and weight were not very outstanding, was able to use his agile posture to obtain a very high movement distance, so he could hit many winning shots with a lower bounce height. The exercise frequency of Participant 3 and Participant 4 was very low, and Participant 3 with below-average distance moved was older, so they did not perform well in the competition. It was worth mentioning that the maximum bounce height of Participant 3 was quite high due to his incorrect use of footwork, thus providing data to support an observing motor control specialist's assumptions.

5. Conclusions

Here, we demonstrated a methodology using deep learning and binocular positioning for undertaking an individualized analysis of footwork in badminton, advocated in an ecological dynamics rationale, revealing key variations in task performance, based on individual participant characteristics. We showed how adopting the use of such a vision-based performance analysis system could support biomechanists and motor control specialists to examine the continuous (re)organization of movement system degrees of freedom (specific body components like the feet) of performers during competition. Such findings may help biomechanists and motor control specialists better support performance preparation and athlete development [3]. At present, the main problem with research on badminton footwork is that it tends to over-rely on subjective observations of motor control specialists and could benefit from a more quantitative analysis of specific body movements. Long-term understanding of how to enrich each athlete's foundational movement skills repertoire, focusing on relevant capacities like balance, agility, footwork, and dynamic movement, could be further enhanced by data from the implementation of such a vision-based, process-oriented analytics system [45]. We presented an automated methodology using deep learning and binocular positioning that can more accurately, efficiently, and economically record the footwork trajectory of individual athletes during competitive performance.

At present, we have integrated this research into a mature intelligent analysis system and applied it to the analysis of badminton players' performance. A cooperation with Shanghai Institute of Physical Education was established to promote the use of the system. The analysis system includes hardware for video capturing and software for trajectory extraction. For implementation, we will first set up two cameras at the designated positions of the stadium to carry out video acquisition and transmission. Using the captured video clips, we will then train the shoe localization model, which is expected to take 2 h. Finally, we will conduct trajectory extraction and analyze the results. The duration of this process depends on the length of the video included in the analysis. For our experiment, it took only 10 min.

Some limitations of the methods merit discussion, including positioning error and matching failure. Multiple factors can contribute to these issues. First, the situation on the real badminton court is ever-changing, and there may be a scenario in which an athlete hits the shuttlecock and causes the opponent's shoes to exit the camera's field of view. Follow-up research should take measures to address this problem by increasing the camera's field of view or classify and deal with the events of shoes beyond the camera's field of view. Second, when the athlete moves too fast and the camera shutter speed is not fast enough, the contours of the shoes will be blurred, which will affect image recognition. Therefore, it is necessary to test with improved camera equipment with higher frame rates to reduce the occurrence of this situation. Third, when two shoes overlap and one shoe covers the other, it is assumed that the coordinate positions of the two shoes are equal, and this will lead to errors in shoe recognition and subsequent trajectory detection and analyses. Fourth, the result of deep learning recognition of the shoes is a bndbox (x_1, y_1, x_2, y_2), which is a rectangular recognition result. Replacing the shoe coordinates with the midpoint of the bottom edge in the view of both cameras will cause errors in binocular positioning, Therefore, finding a more universal and accurate method of approximating points is the key to improving the recognition accuracy. Fifth, the problem of shoe matching failure will directly lead to the error of the final positioning result. Failure to match shoes in the binocular positioning will cause shoe positioning errors, and failure to match shoes in the frame-by-frame matching procedure will lead to shoes classified into the wrong category. Finding a more reliable way to distinguish different shoes is the key to solving shoe matching errors. Sixth, since the current positioning method is only suitable to no more than two shoes in a half court, it can only be used to track the footwork trajectory of a single athlete in either of the half courts. How to adapt our methods to two athletes in the same half court (such as in a doubles game) is the focus of future research. Finally, it would be interesting to apply our methods, in future research, to further identify and study different types of footwork needed in performance on a badminton court, such as the movement footwork performed from deep court to the net or the opposite, as well as from forehand to backhand and vice-versa.

Declarations

Author contribution statement

Jiabei Luo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Yujie Hu: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Keith Davids & Cade Gouin: Analyzed and interpreted the data; Wrote the paper.

Di Zhang: Performed the experiments; Analyzed and interpreted the data.

Xiang Li: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Xianrui Xu: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

Xiang Li was supported by National Natural Science Foundation of China [41771410].

Data availability statement

Data will be made available on request.

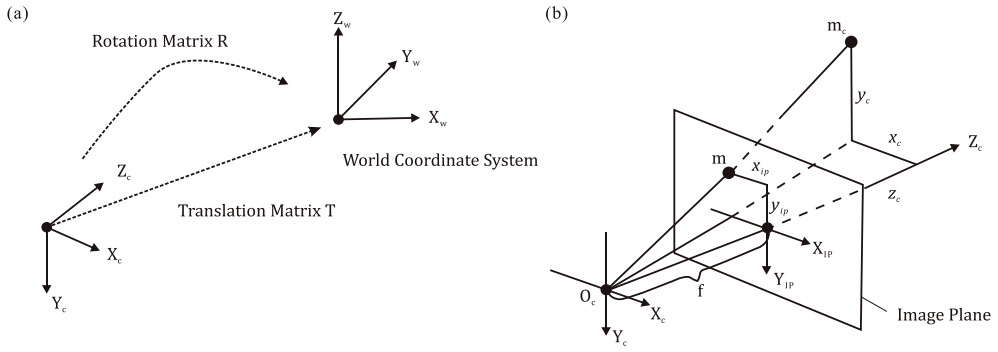
Declaration of interests statement

The authors declare no conflict of interest.

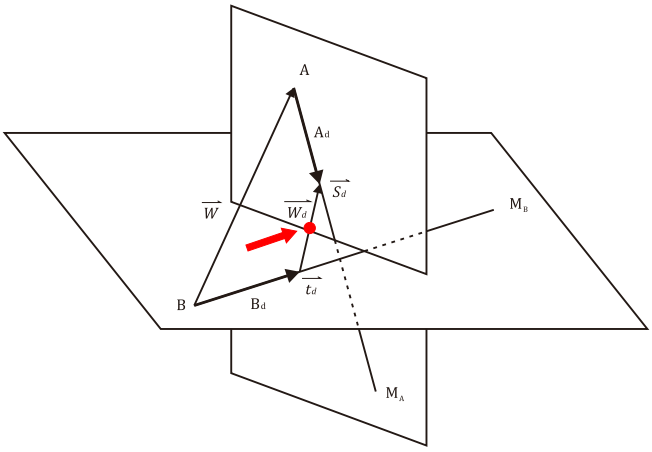
Additional information

No additional information is available for this paper.

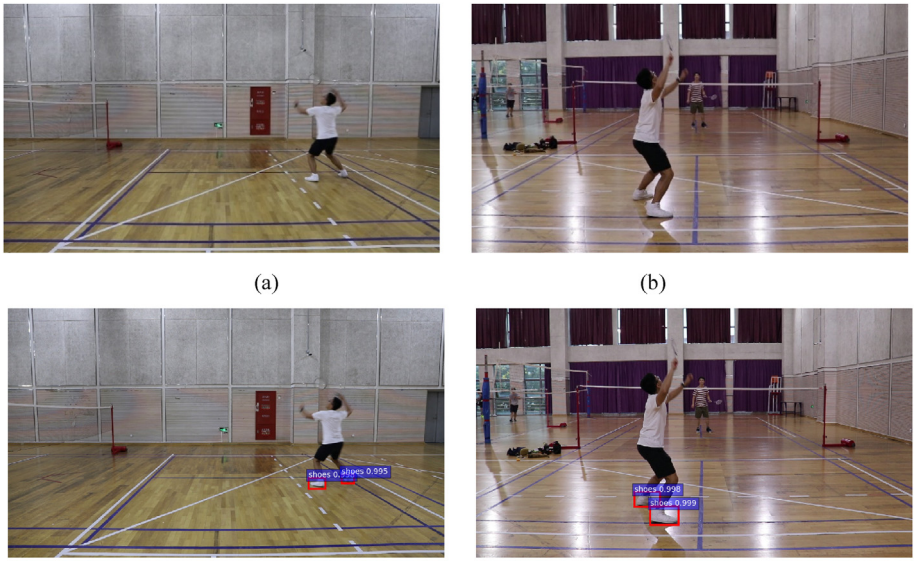
Appendix A. Illustration of (a) rigid body transformation and (b) projection imaging.



Appendix B. Solving the intersection point of two lines.



Appendix C. Sample images captured by cameras A (a) and B (b) at the same time frame and corresponding shoe localization results in (c) and (d).



Appendix D. Accuracy measurement of the binocular positioning method.

Actual world coordinates	Estimated world coordinates using binocular positioning	Error (m)
(2.0,2.0,0.225)	(1.973,1.912,0.203)	0.0946
(3.0,2.0,0.225)	(2.859,2.054,0.212)	0.1515
(4.0,2.0,0.225)	(3.887,1.916,0.251)	0.1432
(5.0,3.0,0.225)	(5.101,3.128,0.206)	0.1642
(5.0,4.0,0.225)	(4.989,4.114,0.189)	0.1201
(5.0,5.0,0.225)	(4.868,4.918,0.212)	0.1559
(5.0,6.0,0.225)	(5.161,5.934,0.221)	0.1740
(4.0,6.0,0.225)	(3.971,5.879,0.251)	0.1271
(3.0,5.0,0.225)	(2.941,4.912,0.219)	0.1061
(3.0,4.0,0.225)	(2.933,4.012,0.244)	0.0707
(3.0,3.0,0.225)	(2.891,2.958,0.253)	0.1201

Appendix E. Characteristics of the study participants

	Sex	Age	Height(m)	Weight (kg)	Exercise frequency
Participant 1	male	22	178	80	1 year
Participant 2	male	28	170	56	3 months
Participant 3	male	46	171	68	1 year
Participant 4	female	21	163	53	3 years
Participant 5	female	22	169	54	1 month
Participant 6	male	23	175	67.5	2-3 weeks

References

[1] K. Davids, G.J. Savelsbergh, S. Bennett, J. Van der Kamp (Eds.), Interceptive Actions in Sport: Information and Movement, 2002.

[2] S. Maftei, Study regarding the specific of badminton footwork, on different levels of performance, Conf. Proc. eLearning Softw. Educ. (eLSE) 3 (1) (2017) 161–166.

[3] R. Valldecabres, J. Richards, A.M. De Benito, The effect of match fatigue in elite badminton players using plantar pressure measurements and the implications to injury mechanisms, Sports BioMech. (2020) 1–18.

[4] W.K. Lam, R. Ding, Y. Qu, Ground reaction forces and knee kinetics during single and repeated badminton lunges, J. Sports Sci. 35 (6) (2017) 587–592.

[5] Y. Hong, S.J. Wang, W.K. Lam, J.T.M. Cheung, Kinetics of badminton lunges in four directions, J. Appl. Biomech. 30 (1) (2014) 113–118.

[6] G. Kuntze, N. Mansfield, W. Sellers, A biomechanical analysis of common lunge tasks in badminton, J. Sports Sci. 28 (2) (2010) 183–191.

[7] Y. Wang, W. Fang, J. Ma, X. Li, A. Zhong, Automatic badminton action recognition using cnn with adaptive feature extraction on sensor data, in: International Conference on Intelligent Computing, Springer, Cham, 2019, pp. 131–143.

[8] S. Ramasinghe, K.M. Chathuramali, R. Rodrigo, Recognition of badminton strokes using dense trajectories, in: 7th International Conference on Information and Automation for Sustainability, IEEE, 2014, pp. 1–6.

[9] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) 1, 2005, pp. 886–893.

[10] D. Araújo, M. Couceiro, L. Seifert, H. Sarmento, K. Davids, Artificial intelligence in sport performance analysis, Routledge, London, 2021.

[11] C.T. Woods, I. McKeown, M. O'Sullivan, S. Robertson, K. Davids, Theory to practice: performance preparation models in contemporary high-level sport guided by an ecological dynamics framework, Sports Med. Open 6 (1) (2020) 1–11.

[12] C. McCosker, F. Otte, M. Rothwell, K. Davids, Principles for technology use in athlete support across the skill level continuum, Int. J. Sports Sci. Coach. 17479541211033471 (2021).

[13] S. Barris, D. Farrow, K. Davids, Increasing functional variability in the preparatory phase of the takeoff improves elite springboard diving performance, Res. Q. Exerc. Sport 85 (2014) 97–106.

[14] C. Caballero, K. Davids, B. Heller, J. Wheat, F. Moreno, Movement variability emerges in gait as adaptation to task constraints in dynamic environments, Gait Post. 70 (2019) 1–5.

[15] M.S. Couceiro, G. Dias, D. Araújo, K. Davids, The ARCANE project: how an ecological dynamics framework can enhance performance assessment and prediction in football, Sports Med. 46 (12) (2016) 1781–1786.

[16] J.M. Giménez-Egido, E. Ortega, I. Verdu-Conesa, A. Cejudo, G. Torres-Luque, Using smart sensors to monitor physical activity and technical–tactical actions in junior tennis players, Int. J. Environ. Res. Publ. Health 17 (3) (2020) 1068.

[17] J. Channells, B. Purcell, R. Barrett, D. James, Determination of rotational kinematics of the lower leg during sprint running using accelerometers, BioMEMS Nanotechnol. II 6036 (2006, January), 603616.

[18] M. Cornacchia, K. Ozcan, Y. Zheng, S. Velipasalar, A survey on activity detection and classification using wearable sensors, IEEE Sensor. J. 17 (2) (2016) 386–403.

[19] S. Barris, C. Button, A review of vision-based motion analysis in sport, Sports Med. 38 (12) (2008) 1025–1043.

[20] S. Nepal, U. Srinivasan, G. Reynolds, Automatic detection of 'Goal' segments in basketball videos, in: Proceedings of the Ninth ACM International Conference on Multimedia, 2001, October, pp. 261–269.

[21] R. Urtasun, D.J. Fleet, P. Fua, Monocular 3D tracking of the golf swing, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2, 2005, June, pp. 932–938.

[22] D. He, L. Li, L. An, Study on sports volleyball tracking technology based on image processing and 3D space matching, IEEE Access 8 (2020) 94258–94267.

[23] K. Host, M. Ivasic-Kos, M. Pobar, Tracking handball players with the DeepSORT algorithm, ICPRAM (2020) 593–599.

[24] T. Guo, K. Tao, Q. Hu, Y. Shen, Detection of ice hockey players and teams via a two-phase cascaded CNN model, IEEE Access 8 (2020) 195062–195073.

[25] J. Ren, J. Orwell, G.A. Jones, M. Xu, Tracking the soccer ball using multiple fixed cameras, Comput. Vis. Image Understand 113 (5) (2009) 633–642.

[26] N.A. Rahmad, N.A.J. Sufri, N.H. Muzamil, M.A. As'ari, Badminton player detection using faster region convolutional neural network, Indonesian J. Elect. Eng. Comp. Sci. 14 (3) (2019) 1330–1335.

[27] C.Z. Shan, E.S.L. Ming, H.A. Rahman, Y.C. Fai, Investigation of upper limb movement during badminton smash, in: 2015 10th Asian Control Conference (ASCC), IEEE, 2015, pp. 1–6.

[28] H. Li, Y.L. Chen, T. Chang, X. Wu, Y. Ou, Y. Xu, Binocular vision positioning for robot grasping, in: 2011 IEEE International Conference on Robotics and Biomimetics, IEEE, 2011, pp. 1522–1527.

[29] H. Shishido, Y. Kameda, I. Kitahara, Y. Ohta, 3D position estimation of badminton shuttle using unsynchronized multiple-view videos, in: Proceedings of the 7th Augmented Human International Conference 2016, 2016, February, pp. 1–2.

[30] C.L. Lee, Badminton Shuttlecock Tracking and 3D Trajectory Estimation from Video, 2016.

[31] Z. Chen, R. Li, C. Ma, X. Li, X. Wang, K. Zeng, 3D vision based fast badminton localization with prediction and error elimination for badminton robot, in: 2016 12th World Congress on Intelligent Control and Automation (WCICA), IEEE, 2016, pp. 3050–3055.

[32] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.

[33] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[34] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[35] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2016) 1137–1149.
[36] L.A. Monezi, A. Calderani Junior, L.A. Mercadante, L.T. Duarte, M.S. Misuta, A video-based framework for automatic 3D localization of multiple basketball players: a combinatorial optimization approach, Front. Bioeng. Biotechnol. 8 (2020) 286.
[37] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.
[38] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
[39] G.E. Hinton, Rectified linear units improve restricted boltzmann machines Vinod Nair, 2010.
[40] S. Jiao, L. Li, H. Qu, M. Zhang, Research on the influence of camera position on reconstruction accuracy in binocular vision, in: Eleventh International Conference on Graphics and Image Processing (ICGIP 2019) 11373, International Society for Optics and Photonics, 2020, January, p. 113732L.
[41] G. Hinton, N. Srivastava, K. Swersky, Neural networks for machine learning, in: Lecture 6a overview of mini-batch gradient descent 14, 2012, p. 2.
[42] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European Conference on Computer Vision, Springer, Cham, 2014, pp. 818–833.
[43] T. Huang, Y. Li, W. Zhu, An auxiliary training method for single-player badminton, in: 2021 16th International Conference on Computer Science & Education (ICCSE), IEEE, 2021, pp. 441–446.
[44] C. Button, L. Seifert, J.-Y. Chow, D. Araújo, K. Davids, Dynamics of Skill Acquisition: an Ecological Dynamics Rationale, second ed., Human Kinetics, Champaign, Ill, 2020.
[45] J.R. Rudd, C. Pesce, B.W. Strafford, K. Davids, Physical literacy-a journey of individual enrichment: an ecological dynamics rationale for enhancing performance and physical activity in all, Front. Psychol. 11 (2020) 1904.