IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 June 2023, accepted 16 August 2023, date of publication 22 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307620

RESEARCH ARTICLE

Skeleton Based Keyframe Detection Framework for Sports Action Analysis: Badminton Smash Case

MUHAMMAD ATIF SARWAR[1], YU-CHEN LIN[1], YOUSEF-AWWAD DARAGHMI[2], TSÌ-UÍ ÍK[1], (Member, IEEE), AND YIH-LANG LI[1], (Member, IEEE)

[1]Department of Computer Science, College of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[2]Department of Computer Systems Engineering, Palestine Technical University–Kadoorie, Tulkarm 310, Palestine

Corresponding author: Tsì-Uí Ík (cwyi@nctu.edu.tw)

This work of Tsì-Uí Ík was supported in part by the Ministry of Science and Technology, Taiwan under grants MOST 110-2627-H-A49-001, MOST 110-2221-E-A49-063-MY3, and MOST 111-2622-E-A49-009. This work was financially supported by the Center for Open Intelligent Connectivity from The Featured Areas Research Center Program within the Higher Education Sprout Project framework by the Ministry of Education (MOE) in Taiwan.

ABSTRACT The analysis of badminton player actions from videos plays a crucial role in improving athletes' performance and generating statistical insights. The complexity and speed of badminton movements pose unique challenges compared to everyday activities. To analyze badminton player actions, we propose a skeleton-based keyframe detection framework for action analysis. Keyframe detection is widely used in video summarization and visual localization due to its computational efficiency and memory optimization compared to analyzing all frames of a video. This framework segments the complex macro-level activity into micro-level segments and analyzes each micro-level activity individually. Firstly, it extracts skeleton data from a motion sequence video using 3D:VIBE pose estimation. Then, the keyframe detection module explores the sequence of activity frames and identifies keyframes for each micro-level activity, including start, ready, strike, and end. Finally, the posture and movement detection modules analyze the posture and movement data to identify specific activities. This framework is implemented in the device called CoachBox. The proposed framework is evaluated using the mean absolute error on a dataset. The average mean absolute error for the keyframe detection module is less than 0.168 seconds, and the striking moment detection has an error of only 0.033 seconds. Additionally, a coordinate transform method is provided to convert body coordinates to real-world coordinates for visualization purposes.

INDEX TERMS Keyframe detection, action analysis, skeleton detection, coordinate transform, action analysis framework.

I. INTRODUCTION

Sports Action Recognition (SAR) is a challenging task used for various sports, including soccer, volleyball, basketball, tennis, and badminton [1], [2], [3], [4]. SAR detects and recognizes actions during competitions, matches, warm-ups, and training sessions [5], [6]. SAR-based applications have been extensively utilized by sports analysts and coaches to enhance athletes' performance. The main objective of SAR applications is to identify the athlete's actions from an unknown video sequence, determine the action's duration and type, monitor a player's performance, track their movements, recognize the performed action, compare various actions, compare different kinds and skills of performances, or perform automatic statistical analysis [7], [8], [9].

Badminton is a highly technical sport that can greatly benefit from SAR-based applications for analyzing player actions. Recently, research on badminton actions [10], [11] has made rapid progress in monitoring athletes' performance. It involves comparing various actions performed

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

by different players or multiple executions by a single player, such as smash, clear, and backhand. These advancements assist in practicing techniques and improving playing styles [12], [13]. Ramasinghe et al. [14] proposed an accurate approach for badminton stroke recognition using RGB-D data. They utilized dense trajectories and trajectory-aligned HOG features to classify four stroke classes, including smash, forehand, backhand, and break, using an SVM classifier. Wang et al. [15] inserted a chip into the badminton racket to collect data, which was then analyzed by a deep convolutional neural network (CNN) for action recognition. Another example is PitchAI, [16] a mobile app that analyzes pitching movements using neural networks and 3D skeleton data to calculate movement features and evaluate the kinetic chain. The CoachBox [17] is a computer vision-based system designed to monitor badminton strike actions and improve player performance.

Intelligent badminton action recognition techniques [10], [11] have been developed to provide objective analysis and evaluation of badminton, leading to improved accuracy in performance analysis and more efficient training programs [18]. However, assessing and objectively measuring badminton activity from the macro-level to the micro-level is more challenging compared to daily activities. These techniques do not provide detailed analysis of activities at the micro-level, such as specific actions, movements, or behaviors of individuals or objects. They also lack the ability to capture fine-grained aspects of an activity and may overlook important seconds-level details or specific interactions. For instance, badminton strokes, e.g. the smash, can be further segmented into micro-level activity to enhance the analysis of the player's complete action. This level of granularity is necessary to capture and understand the nuances of each attribute of micro-level badminton activity. Therefore, there is a need for a system that can comprehensively analyze badminton activity from the macro-level to the micro-level, providing insights into each attribute of micro-level badminton activity.

In order to perform badminton video analysis and process activity from the macro-level to the micro-level, it is crucial to extract relevant and essential information about the badminton player. One important step in video analysis is the extraction of keyframes, as they contain significant information that provides a summary of the entire video sequence. In this paper, we propose a novel framework that utilizes badminton videos to extract keyframes from motion sequences. This framework combines pose estimation techniques and a keyframe detection algorithm to classify activities from the macro-level to the micro-level, as illustrated in Fig. 1. The recognition of skeleton-based activity is achieved by analyzing the sequence of skeleton keypoints over time using 3D:VIBE [19] from RGB data [20], [21], [22]. The keyframe detection module extracts key-pose frames from a series of activity frames, while the pose and movement modules detect and recognize key poses based on predefined rules. By integrating these components, our framework enables the analysis of badminton videos at various levels,

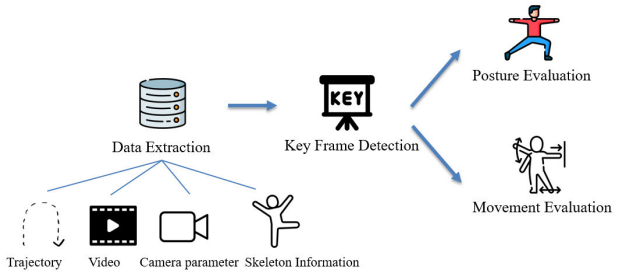


FIGURE 1. Skeleton based keyframe detection and action recognition framework.

providing insights into both the macro-level and micro-level activities performed by the players.

The proposed framework is implemented in CoachBox, a stereo vision device equipped with two cameras that automatically captures badminton actions for learning purposes. The framework utilizes a database collected from multiple athletes' action data, including badminton smashes, and consists of videos of a total of 600 badminton rallies. To evaluate the proposed methodology, the mean absolute error is used as the performance metric, calculated for each player. The average mean absolute error for the keyframe detection module is less than 0.168 seconds, indicating a high accuracy in detecting keyframes. Furthermore, the striking moment is measured with an average mean absolute error of only 0.033 seconds, demonstrating precise detection of the precise moment of impact. The small magnitude of changes in activity within a short time frame suggests that the proposed framework performs well in capturing and analyzing badminton actions, ensuring that the detected keyframes and striking moments are within an acceptable range.

The primary contributions of this work include the development of the framework, the utilization of a comprehensive dataset, and the evaluation of the methodology using mean absolute error, highlighting the effectiveness and accuracy of the proposed approach are:

- Through our experiments, we demonstrate how incorporating the skeleton-based keyframe selection module assists in achieving effective keyframe features for the badminton activity framework.
- The developed framework extracts skeleton information from the video and identifies keyframe postures for action recognition, reducing the computational resources required compared to using all frames of a video.
- The proposed network pathways are designed to provide real-time feedback for coaches and athletes, enabling them to improve their actions in a timely manner.

The rest of the paper is organized as follows: Section II discusses the related work about badminton activity recognition. Section III presents the methodology and prototype design. Section IV describes the performance evaluation and implementation environment, and section V concludes this research.

II. RELATED METHODS

In this section, we will explain the related methods that are employed to recognize badminton shot actions. The stereo vision cameras capture the players' video that assists in reconstructing the 3D representation of the player and calculating the court size. Subsequently, OpenPose and VIBE are utilized to extract the 2D and 3D skeletons from the video. Finally, a keyframe detection module is employed to extract key frames for sub-action content analysis. Additionally, the details of the badminton action recognition and keyframe extraction methods will be provided.

A. STEREO VISION CAMERA

We utilized two stereo vision cameras with different viewing angles to accurately capture the depth information of badminton actions by employing the principle of triangulation [23]. By obtaining depth information from stereo vision, it becomes possible to reconstruct a three-dimensional (3D) representation of the scene. Prior to calculating the 3D positions, it is necessary to determine the intrinsic and extrinsic matrices for each camera.

To obtain the intrinsic matrix and distortion parameters for each camera, a checkerboard pattern was used to assist in calibration. Additionally, the extrinsic matrix for each camera in the court was calculated by employing homography mapping from the white field lines of the court to the known court size. Once the camera parameters were obtained, the 3D points were triangulated from a set of points calculated using two different perspective images.

B. HUMAN SKELETON DETECTION

Human skeleton detection can be broadly categorized into two types: 2D skeleton prediction models and 3D skeleton prediction models.

1) 2D HUMAN SKELETON DETECTION

Firstly, we utilize OpenPose [20], which is an open-source state-of-the-art method based on Part Affinity Fields (PAF) to track human pose in badminton court. PAF provides vectors that connect one joint point to the next, capturing the relationships between different body parts. OpenPose is a highly capable framework that enables the detection and tracking of multiple people's poses simultaneously. This multi-person tracking capability is particularly important as it allows us to account for human interactions and analyze them in natural settings. By leveraging OpenPose, we can accurately track and analyze the poses of badminton player and understand their interactions with each other that aids in visualizing from real-world court to the virtual world.

2) 3D HUMAN SKELETON DETECTION

We utilized the VIBE (Video Inference for Human Body Pose and Shape Estimation) network [19] to extract the 3D human skeleton from a monocular RGB video [24]. The primary objective of VIBE is to accurately estimate the 3D body pose and shape of badminton players from such videos. VIBE adopts an end-to-end architecture, transforming 2D input images to 3D skeleton coordinates through a generative adversarial network (GAN) [25]. To capture the temporal relationship of the video frames and enhance action coherence, VIBE incorporates a gated recurrent unit (GRU) [26].

To train VIBE, a mixed dataset comprising 2D and 3D data from MPI-INF-3DHP [27], Human3.6M [28], and 3DPW [29] is employed. This diverse dataset ensures robust training and generalization of the network. VIBE's performance is evaluated using the Percentage of Correct Keypoints (PCK) metric, achieving an impressive correctness score of 89.3%.

VIBE detects 49 key points of the skeleton, providing a detailed understanding of the actions captured in the badminton video. This comprehensive set of key points enables a thorough analysis of the player action content, facilitating further interpretation.

C. TRAJECTORY DETECTION

We employed the TrackNetv2 network, as described in [30], to track shuttlecocks and visualize their positions in the virtual world court. TrackNetV2 is specifically designed to excel in detecting small, fast-moving objects such as shuttlecocks in video footage. It operates on a frame-by-frame basis, accurately determining the shuttlecock's position in each frame.

The architecture of TrackNetV2 follows an encoder-decoder structure. The encoder acts as a feature extractor, utilizing convolutional kernels to capture image clues and condensing the features through max-pooling operations. Conversely, the decoder expands the feature maps to generate the prediction function, enabling accurate shuttlecock tracking.

TrackNetV2 is trained on dataset that contains 55563 frames including 15 broadcast videos of professional games and 3 amateur games. In order to prevent overfitting, we collected an additional 125 rally videos with diverse backgrounds and filming angles. Approximately 2,500 to 3,000 frames were included from each video. TrackNetV2 accuracy respectively reach to 98.7% in the training phase and 85.4% in a test on a new match. Moreover, TrackNetV2 exhibits a processing speed of 31.84 frames per second (FPS), which greatly facilitates shuttlecock tracking in our approach.

D. KEYFRAME EXTRACTION

Several researchers have proposed various keyframe extraction methods using different strategies. Phan et al. [31] introduced an efficient framework named KFSENet for action recognition in videos, incorporating keyframe extraction based on skeleton deep learning architectures. Kim et al. [32] proposed a bidirectional consecutively connected two pathway network (BCCN) for efficient gesture recognition using a Skeleton-Based Keyframe Selection Module. Lv et al. [33] developed a sports action classification system for accurately classifying athletes' actions based on keyframe extraction.

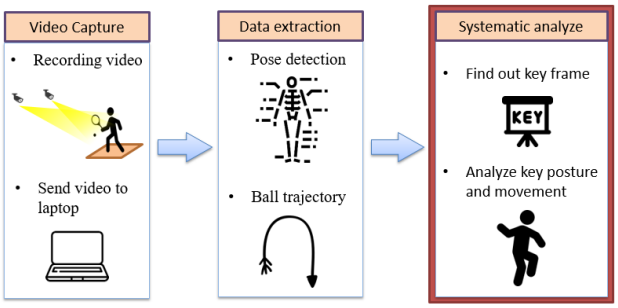


FIGURE 2. Keyframe based activity recognition methodology.

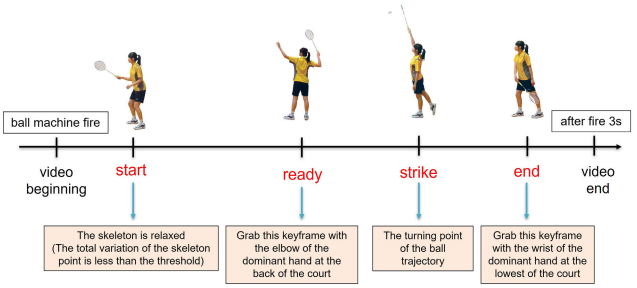


FIGURE 3. Keyframe detection module.

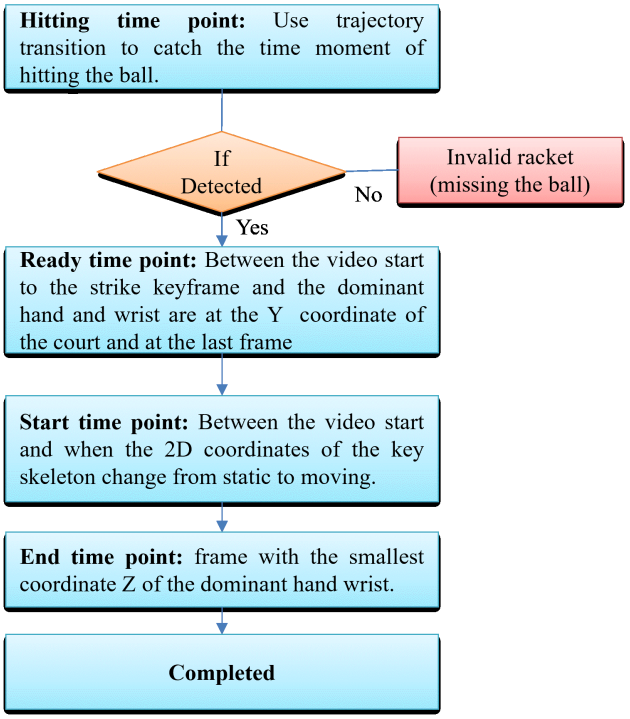


FIGURE 4. Working flow of keyframe detection module.

III. KEYFRAME BASED ACTION ANALYSIS METHODOLOGY

keyframe based action analysis framework contains four main modules including data extraction, keyframe detection, posture detection and movement detection, as illustrated in Fig. 2.

A. DATA EXTRACTION MODULE

The first module of the system is data extraction, which consists of five steps: multi-view video, camera parameters, skeleton detection, trajectory detection, and coordinate system. The multi-view video system is equipped with two stereo-vision cameras to capture the game video of badminton players. The camera parameters, including the intrinsic and extrinsic matrix and distortion parameters, are calculated using the methodology defined in subsection II-A of section II. The skeleton detection utilizes the 3D:VIBE [19] method to identify the keypoints of the human body from the video. Similarly, the TrackNetV2 [30] network tracks the 3D position of the badminton shuttle from the video.

After the calculation of the above steps, the coordinate system incorporates real-world court coordinates synchronized with timestamps and camera parameters. The court coordinates are defined with the court as the origin, the short side as the X-axis, the long side as the Y-axis, and the Z-axis pointing upward towards the ground's normal vector. The corrected camera parameters are utilized to establish the relative relationship between the two cameras and the court origin, enabling the visualization of the players' actions and the ball. Lastly, the body and shuttle coordinates extraction are mapped onto the world coordinate system to be visualized in the virtual court.

B. KEYFRAME DETECTION MODULE

Keyframe detection is utilized to extract key-pose frames from a series of action frames. In a sequence of action frames, there are several key poses that specifically represent certain actions. Our proposed method for action recognition extracts key pose frames from videos instead of analyzing the entire video sequence frame-by-frame. This approach significantly reduces the amount of data that needs to be processed, as keyframes only capture the most significant parts of the video where noticeable changes occur.

To identify these keyframes, we focus on studying the badminton smash action, which can be roughly divided into four key poses. These poses are determined based on the skeleton keypoints and ball trajectory, as illustrated in Figure 3. By analyzing the skeleton and trajectory information, each smashing video is segmented into several one-shot videos using the trajectory turning point as a reference. The one-shot videos are synchronized with timestamps from two different perspectives, and the entire action can be segmented based on four key posture positions, including the start, ready, strike, and end poses.

First, the ball trajectory turning point is calculated to determine the moment of the strike position. The turning point can be calculated by analyzing the variation product of the Y vector, considering that badminton is played along the long axis. If the trajectory turning point is not found, it indicates that the ball was not hit, and the racquet is considered invalid. This process is illustrated in the flow chart shown in Figure 4.

Second, the ready posture keyframe is determined based on the position of the elbow as the back-most point of the court from the beginning of the video up to the keyframe of the strike. The elbow keypoint plays a crucial role in identifying the ready posture in the badminton smash action.

$$\sum_{f=f_0}^{f_0+k*fps} \sum_{i=0}^{w} |j_{f+1}^i - j_f^i| < const \tag{1}$$

Third, the start posture keyframe is identified as the frame with the least skeleton variation. The start posture is determined based on the relaxed position of the body, either considering the entire body or specific joints in a relaxed position. The relaxation position is calculated using the Euclidean distance equation [34], as shown in Equation 1. In Equation 1, f_0 represents the initial frame, k is the time in seconds, fps is the camera frame rate per second, w denotes a set of important skeleton points (such as the shoulder, elbow, neck, wrist), and j^i represents the 3D coordinates of skeleton point i in the court. The equation implies that the sum of the moving distances of all skeleton points w within k seconds should be less than a threshold. The threshold can be determined by calculating the sum of the first quarter of the movement of all skeleton keypoints (e.g., if the keypoints' movement is 100cm, a threshold value of 25cm is considered). If the Euclidean distance variation for the w skeleton points within k seconds is less than the threshold, it is considered as a start posture keyframe for the smash action.

Finally, the end posture keyframe is determined by considering the position of the wrist as the lowest point on the court from the keyframe of the strike to the end of the video. The timestamp can also be utilized to assist in identifying the keyframe, especially when the keyframe is located very close to the end of the video. The overall analytical framework for the badminton smash action, which includes action description, keyframe detection, posture analysis, and movement analysis, is defined in Table 2.

Furthermore, this framework can also be applied to analyze other ball actions in badminton, such as the forehand/backhand high ball and the forehand/backhand cut ball. The process involves specifying keyframe restrictions and extracting the features of each posture and movement for analysis. The only differences lie in the number of keyframes and the specific action judgments. For instance, the analysis of the hit angle is required for high ball, smash ball, and cut ball, but the landing location of the ball and the ball speed requirements may vary. Therefore, the judgments for different types of actions need to be tailored accordingly. In this case, the number of keyframes remains the same for these three actions, but it will differ from the analysis of a flat ball.

C. POSTURE AND MOVEMENT DETECTION MODULE

The posture and movement evaluation modules analyze both the static and dynamic movements of the action features. Once the keyframes are detected, the entire action is segmented into several segments. The static pose is defined by the stillness captured in the keyframe, while the movement occurring near the keyframes or between two keyframes represents the dynamic pose. For instance, in the case of a badminton smash, there are four key postures: start, ready, hit, and end. Additionally, there are three key movements between adjacent keyframes, including the preparation period, swing period, and closing period. The extracted features from these key postures and movements are valuable as they can be evaluated based on expert feedback. This evaluation process aids both beginners and professional players in assessing their actions.

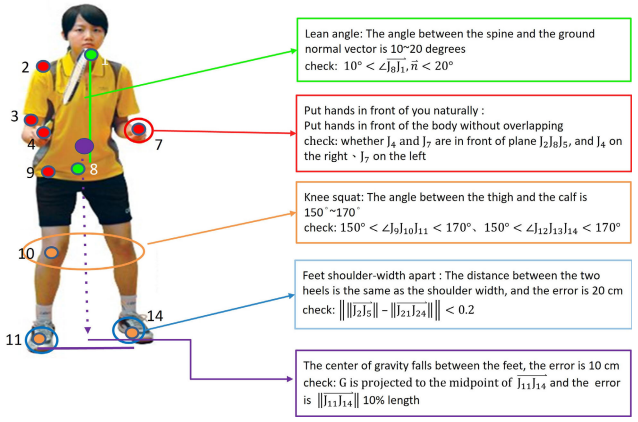


FIGURE 5. Posture detection mechanism for badminton smash start post.

The posture detection module measures the position of each keyframe posture, including start, ready, strike, and end, based on rules defined by coaches and experts [16]. For example, the start posture of a badminton smash is defined by five skeleton keypoints positions. First, the lean angle, which is the angle between the spine vector $\overline{J_1 J_8}$ and the ground normal vector $\bar{n}$, should be between 10°∼20°. Second, the hands should be placed naturally in front of the body without overlapping, ensuring that the right hand J_4 and left hand J_7 are on their respective sides and facing the plane defined by keypoints $J_2 J_8 J_5$.

Third, the knees keypoints J_{10}, J_{13} should be in a squat position, and the angle between the both thigh keypoints J_9, J_{11} and the both calf keypoints J_{12}, J_{14} should be between 150°∼170°, as shown in the orange rectangle in Fig. 5. Fourth, the absolute distance between the two heel keypoints J_{21} and J_{24} should be the same as the width between the shoulder keypoints (J_2 and J_5), with an acceptable error of about 20 cm. Finally, the center of gravity should fall between the feet, with an acceptable error of about 10 cm, and the gravity vector G should be projected onto the midpoint keypoint (ankle joint keypoints J_{11} and J_{14}) with an absolute vector error of 10%. Similar patterns are used to measure the remaining keyframe postures, including ready, strike, and end.

The movement detection module detects the player movement duration between the adjacent keyframes including the

TABLE 1. Characteristics of the study participants.

Action Description	keyframe detection	Posture Analysis	Movement Analysis
Start: Prepare for the action, relax your body naturally and keep still waiting for the ball.	The skeleton is in relax position (The total variation of the skeleton point is less than the threshold	1: Knee squat 2: Feet shoulder-width apart 3: The center of gravity is between the feet 4: Put your hands in front of you naturally 5: Leaning forward	No Movement
Ready: Stand up with both hands and wait for the ball to hit the right position.	Grab this keyframe with the elbow of the dominant hand at the back of the court.	1: Two arms parallel 2: Dominant foot behind 3: The non-dominant elbow is higher than the shoulder 4: Gravity point on the back	Preparation period: The process from start to ready. The dominant hand pulls back.
Strike: Arm swing to hit the ball	The turning point of the ball trajectory.	1: Hit angle (45) 2: Dominant elbow angle 3: Ball angle 4: Gravity point on the front	Swing period: The process from ready to strike. Use the upper arm with rotation of waist and forearm to hit the ball .
End: Complete the action, the wrist swing to the non-dominant side of the body.	Grab this keyframe with the wrist of the dominant hand at the lowest of the court .	1: Dominant hand wrist near left waist 2: dominant foot forward 3: Ball angle 4: Gravity point on the front	Closing period: The process from strike to end. Closing action after hitting the ball .

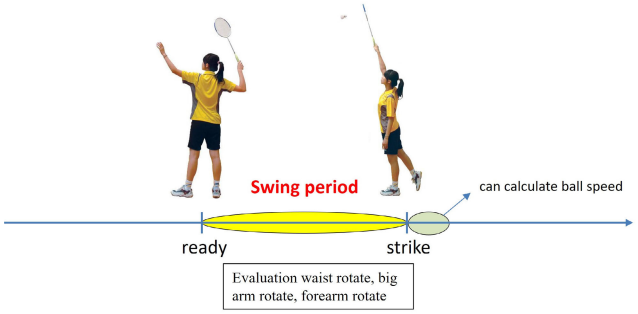
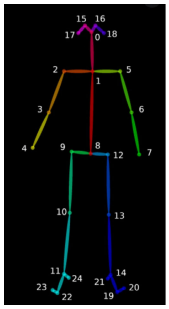


FIGURE 6. Movement detection: smash swing period.



Segment	Males	Females	
Head	8.26	8.20	$(J_{17}+J_{18})/2$
Trunk	46.84	45.00	$(J_2+J_5+J_9+J_{12})/4$
Upper Arm	3.25	2.90	$(J_2+J_3+J_5+J_6)/4$
Forearm	1.87	1.57	$(J_3+J_4+J_6+J_7)/4$
Hand	0.65	0.50	$(J_4+J_7)/2$
Thigh	10.50	11.75	$(J_9+J_{10}+J_{12}+J_{13})/4$
Lower Leg	4.75	5.35	$(J_{10}+J_{11}+J_{13}+J_{14})/4$
Foot	1.43	1.33	$(J_{11}+J_4+J_6+J_7)/4$

FIGURE 7. OpenPose keypoints based gravity center.

preparation period, the swing period, and the closing period. The movement detection mainly focuses on changes and accumulation in joints keypoints including rotation, action sequence, kinetic chain, ball speed, etc. A player movement is detected based on the movement features which is roughly divided into four common categories including (i) angle; (ii) spatial comparison; (iii) rotation; (iv) relax positions. All these angles, lines and planar are composed of human joint keypoints, ball positions, coordinate points, or gravity points of the human body. For example, the swing period is detected between the ready and strike posture as illustrated in Fig. 6. The swing period is measured based on the movement of upper arm with rotation of the waist and the forearm to hit the ball. Similarly, the rest of movement period is determined based on the rules shown in table 2 in the movement analysis column. The posture and movement detection modules assists in determining the player's action, and an action is determined based on the body movement position either the whole body joints or some joints.

1) GRAVITY POINT

It is worth mentioning that the center of gravity plays a crucial role in ensuring the correctness of the player's action. The change in the center of gravity helps beginner players gain a clearer understanding of their actions and improve them, as illustrated in Fig.7. The center of gravity is a significant aspect of biomechanics and locomotion, aiding in the modeling of the human body and its activities. It is instrumental in assessing static positions and various types of movement techniques. The center of gravity [35] of the whole body is calculated using the weighted average of body keypoints, as shown in equation2, where w_i represents the weight average of keypoints and j_i denotes keypoint i of the body.

$$G = \frac{\sum w_i * j_i}{\sum w_i} \tag{2}$$

2) COORDINATE TRANSFORM

The coordinate transform method is used to map the human body coordinates (kp_{pose}), generated by VIBE, onto the virtual court coordinate system (kp_{court}), as depicted in Fig. 8. In the figure, the red dotted rectangle represents the camera coordinates, and the blue color notations represent unknown parameters. The intrinsic matrix and extrinsic matrix are represented by K_s, which is obtained using the perspective projection formula of computer vision and perspective-n-Point pose computation [36]. R_{pose} and T_{pose} are calculated

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & u_o \\ 0 & \alpha_y & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\mathrm{kp_{img}} = \mathbf{Ks} * [\mathbf{R_{pose}} \quad \mathbf{T_{pose}}] * \mathrm{kp_{pose}} \rightarrow \text{Camera coordinate system}$$

$$= [\mathbf{R_{court}} \quad \mathbf{T_{court}}] * \mathrm{kp_{court}} \rightarrow \text{Camera coordinate system}$$

The blue color parameter is needed to calculate

Step1: Use solvePnP to calculate $\mathbf{R_{pose}}$ and $\mathbf{T_{pose}}$

Step2: $\mathrm{kp_{court}} = \mathbf{R_{court}}^{-1} * (\mathbf{R_{pose}} * \mathrm{kp_{pose}} + \mathbf{T_{pose}} - \mathbf{T_{court}})$

FIGURE 8. Coordinate transform method.

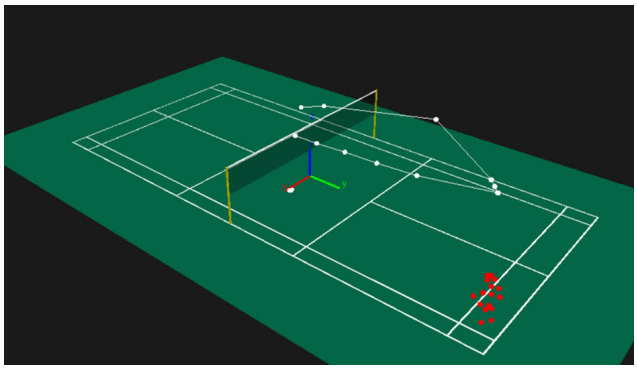


FIGURE 9. Visualization of coordinates points with ball trajectory in virtual court.

by applying rotation and taking the transpose of the skeleton pose keypoints. After rotation and transpose of the skeleton pose, a dot product is performed with the body coordinate kp_{pose} to obtain the body keypoints image kp_{img}, which visualizes the player's body keypoints in the virtual court.

Similarly, R_{court} and T_{court} represent the court coordinates rotation and transpose notations. These notations, along with the dot product of kp_{court}, also help in obtaining the body keypoints image kp_{img} to visualize the player's body keypoints in the virtual court. The notation kp_{court} is obtained using the formula defined in step 2. After finding the body and court keypoints, they are visualized in the virtual court along with the shuttle trajectory, as illustrated in Fig. 9, where white points indicate the badminton trajectory and red points represent the body keypoints.

IV. PERFORMANCE EVALUATION AND IMPLEMENTATION ENVIRONMENT

A. DATASET

The badminton games dataset was captured independently on multiple subjects and used for performance evaluation. This dataset has synchronized multi-view videos and labeled keyframes that are defined by each evaluation algorithm. The dataset collection process include (i) fixing the position and angle of the two cameras, (ii) the testing player needs to stand in the red rectangle of the two pictures in CoachBox like the fig 10, (iii) a ball machine also needs to be placed in a fixed

FIGURE 10. Camera view angle and tester standing position.

position so that tester can hit the ball better, (iv) selecting the court line corner to calculate the extrinsic matrix, (v) starting the test.

The dataset contains six types of ball for each of the ten people that have different gender, ages, and levels, as stated in table 2. The ball types include forehand and backhand smash, forehand and backhand high ball, forehand and backhand cut ball, and involve a total of six actions of 10 people. Each action is done 10 times, which means the ball machine will serve 10 consecutive balls as the data collection. So, a short video of a total of 600 rallies that are pre-edited and stored in the dataset. The intrinsic matrix and extrinsic matrix are saved in the dataset. The collecting process is that the player arrives at the designated position on the court, and then another person presses the start test button. The ball machine first serve two balls for initial testing, and these two balls will not be included in the evaluation. Then, the official test starts by serving 10 consecutive balls for the data collection and for changing the next action.

B. PERFORMANCE EVALUATION

1) KEYFRAME DETECTION EVALUATION

The keyframe detection module is evaluated using the above mentioned dataset based on the mean absolute error which compares the ground truth and prediction frame labeling. The keyframe module is evaluated according to different categories of players. The first category represents the professional player which is only one in our dataset and the rest belongs to the beginner category. First, we evaluated the module for the professional player and calculated the error. The module used 60 different types of action to detect the four different keyframes postures including the start, ready, strike and end. The average mean absolute errors of the four keyframe posture are shown in Table 3.

The performance of the keyframe detection module is also evaluated on 5 beginner-level players which are randomly decided and their actions frames are 300 actions in total. The average mean absolute errors of the four keyframe posture are shown in Table 4. The results show that the mean absolute error is larger for the beginner players that the professional

TABLE 2. Characteristics of the study participants.

Subject	Sex	Age	Height (cm)	Weight (kg)	Exercise frequency
Participant 1 (P)	male	22	173	67	1 year
Participant 2	male	28	172	70	8 month
Participant 3	male	25	178	74	5 month
Participant 4	male	24	170	71	6 month
Participant 5	male	23	175	76	4 month
Participant 6	female	26	163	53	3 month
Participant 7	female	25	170	55	2 month
Participant 8	female	26	172	59	1 month
Participant 9	female	22	172	61	6 month
Participant 10	female	21	168	60	5 month

TABLE 3. keyframe evaluation on pro-level player.

	T_Start	T_Ready	T_Strike	T_End
MAE(s)	0.168	0.179	0.033	0.164

TABLE 4. keyframe evaluation on beginner-level player.

	T_Start	T_Ready	T_Strike	T_End
MAE(s)	0.238	0.233	0.053	0.500

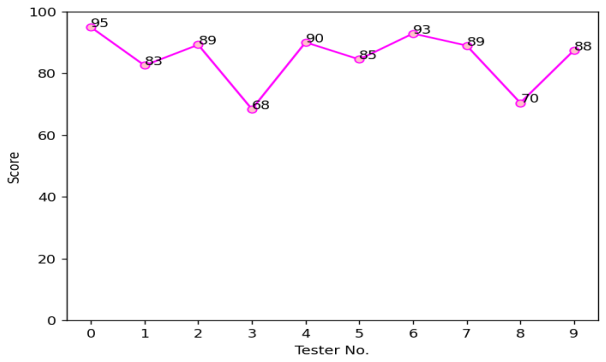


FIGURE 11. Visualization of Athletes performance.

player due to the non-standard action. The reason for the error is generally that the entire action is not completed or the player does not return to the original position.

2) POSTURE EVALUATION

Each posture position is evaluated based on the pre-defined rules in Table 2 and Fig. 5. The athlete who has the most accurate posture position gets the highest score, as illustrated in Fig. 11. The figure shows that the average score of the beginning-level athlete is lower than professional player. The reason is that their actions are inaccurate and incomplete compared to the professional player.

The proposed method is compared with a similar method as shown in Table5. These methods are able to recognize various player actions, such as forehand, backhand, serve, and volley, by analyzing the video frames. The performance of these proposed methods are evaluated on their own datasets, and the

TABLE 5. Comparative analysis with related methods.

Method	accuracy	Number of stroke classes
Miyamori et al. [38]	97.42%	5
Zhu et al. [39]	90.21%	2
Ramasinghe et al.[14]	93.34%	4
Ours	**98.01%**	6

TABLE 6. Number of evaluated strokes of each class.

Stroke class	Number of tested shots	accuracy
Smash	100	98.56%
Forehand	100	**96.27%**
Backhand	100	97.7%
Other	300	**94.01%**

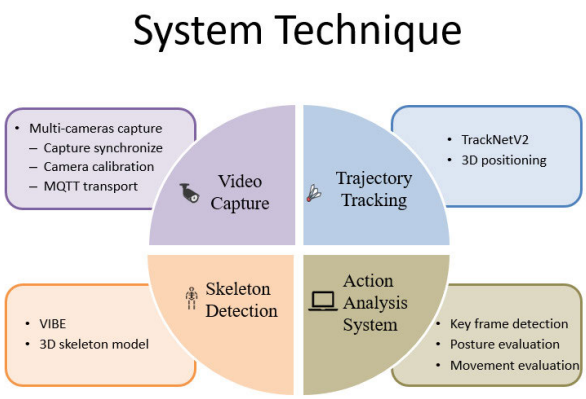


FIGURE 12. CoachBox technology overview.

results showed that our method outperformed the other state-of-the-art methods for player action recognition in badminton videos. Similarly, the proposed method is evaluated on each class to check the performance. Table 6 shows each class result with their respective shots data.

C. CoachBox: SYSTEM TECHNIQUE

The CoachBox's entire system technology is illustrated in Fig. 12. It is divided into four main parts, video capture, shuttlecock trajectory tracking, skeleton detection, and action analysis. The video capture part includes how to synchronize

and record the multi-view video, camera calibration that calculates the intrinsic matrix and the extrinsic matrix, and also includes using MQTT to transport the data from two cameras. Trajectory tracking includes using TrackNetV2 to detect shuttlecock tracks, 3D positioning, trajectory smoothing, etc. Skeleton detection includes using VIBE to detect 3D human skeletons, parsing the output of VIBE, and transforming 3D skeleton points into the court coordinate system. The last part is the action analysis system which uses the framework proposed in this study to systematically analyze the actions, including data extraction, keyframe detection, posture evaluation, and movement evaluation.

V. CONCLUSION

This paper presents a badminton action analysis framework that offers a solution for analyzing and evaluating complex shots, such as the smash, in badminton videos using the keyframe detection module. The framework can handle real-time inputs from badminton games and provides a comprehensive analysis of badminton activity, ranging from macro-level to micro-level analysis, allowing for insights into each attribute of micro-level badminton activity. Furthermore, the framework is implemented on CoachBox, enabling the mapping of player actions and shuttle trajectories onto real-world courts for visualization. This system assists coaches and players in generating analysis reports that provide insights into their games, helping them correct their action poses and reduce the risk of sports injuries. The future work will focus on developing the action description language to translate the coach's defined feature judgments, thus enhancing the algorithm's efficiency and facilitating the systematic integration of all action features.

REFERENCES

[1] K. Host and M. Ivašić-Kos, "An overview of human action recognition in sports based on computer vision," *Heliyon*, vol. 8, no. 6, Jun. 2022, Art. no. e09633.

[2] B. Li and M. Tian, "Volleyball movement standardization recognition model based on convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2023, pp. 1–9, Jan. 2023.

[3] Y. Li, Y. Liu, R. Yu, H. Zong, and W. Xie, "Dual attention based spatial–temporal inference network for volleyball group activity recognition," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 15515–15533, Apr. 2023.

[4] M. Ibh, S. Grasshof, D. Witzner, and P. Madeleine, "TemPose: A new skeleton-based transformer model designed for fine-grained motion recognition in badminton," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 5198–5207.

[5] K. Davids, S. Bennett, G. J. Savelsbergh, and J. Van der Kamp, *Interceptive Actions in Sport: Information and Movement*, 2002.

[6] Ş. Maftei, "Study regarding the specific of badminton footwork, on different levels of performance," in *Proc. eLearning Softw. Educ. (eLSE)*, vol. 13, no. 1. Carol I National Defence Univ. Publishing House, 2017, pp. 161–166.

[7] S.-H. Cheng, M. A. Sarwar, Y.-A. Daraghmi, T.-U. Ik, and Y.-L. Li, "Periodic physical activity information segmentation, counting and recognition from video," *IEEE Access*, vol. 11, pp. 23019–23031, 2023.

[8] K. Soomro and A. R. Zamir, "Action recognition in realistic sports videos," in *Computer Vision in Sports*. Cham, Switzerland: Springer, 2015, pp. 181–208.

[9] S. Zhou, "A survey of pet action recognition with action recommendation based on HAR," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol. (WI-IAT)*, Nov. 2022, pp. 765–770.

[10] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3551–3558.

[11] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[12] W. Liu and J. Ke, "A brief analysis of multi-ball training in badminton," *Educ. Res. Frontier*, vol. 10, no. 4, 2020.

[13] T. Huang, Y. Li, and W. Zhu, "An auxiliary training method for single-player badminton," in *Proc. 16th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2021, pp. 441–446.

[14] S. Ramasinghe, K. G. M. Chathuramali, and R. Rodrigo, "Recognition of badminton strokes using dense trajectories," in *Proc. 7th Int. Conf. Inf. Autom. Sustainability*, Dec. 2014, pp. 1–6.

[15] Y. Wang, W. Fang, J. Ma, X. Li, and A. Zhong, "Automatic badminton action recognition using CNN with adaptive feature extraction on sensor data," in *Proc. 15th Int. Conf. Intell. Comput. Theories Appl. (ICIC)*, Nanchang, China. Cham, Switzerland: Springer, Aug. 2019, pp. 131–143.

[16] Proplayai. *Pitchai*. [Online]. Available: https://proplayai.com/pitchai/

[17] P.-Y. Kuo. *Badminton Smash Visualization System*. [Online]. Available: https://etd.lib.nctu.edu.tw/cgi-bin/gs32/tugsweb.cgi?o=dnctucdr&s=id=%22GT0706568120%22.&searchmode=basic

[18] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal multiplier networks for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7445–7454.

[19] M. Kocabas, N. Athanasiou, and M. J. Black, "VIBE: Video inference for human body pose and shape estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5252–5262.

[20] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1302–1310.

[21] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 466–481.

[22] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Aug. 2017.

[23] W. Luo, Y. Qin, Q. Li, D. Zhang, and L. Li, "Automatic mileage positioning for road inspection using binocular stereo vision system and global navigation satellite system," *Autom. Construction*, vol. 146, Feb. 2023, Art. no. 104705.

[24] W. Liu, Q. Bao, Y. Sun, and T. Mei, "Recent advances of monocular 2D and 3D human pose estimation: A deep learning perspective," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–41, Apr. 2023.

[25] P. Bhattacharjee and S. Das, "Temporal coherency based criteria for predicting video frames using deep multi-stage generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[26] A. Sen and K. Deb, "Categorization of actions in soccer videos using a combination of transfer learning and gated recurrent unit," *ICT Exp.*, vol. 8, no. 1, pp. 65–71, Mar. 2022.

[27] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "Monocular 3D human pose estimation in the wild using improved CNN supervision," in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2017, pp. 506–516.

[28] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.

[29] T. Von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, "Recovering accurate 3D human pose in the wild using IMUs and a moving camera," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 601–617.

[30] N.-E. Sun, Y.-C. Lin, S.-P. Chuang, T.-H. Hsu, D.-R. Yu, H.-Y. Chung, and T.-U. Ik, "TrackNetV2: Efficient shuttlecock tracking network," in *Proc. Int. Conf. Pervasive Artif. Intell. (ICPAI)*, Dec. 2020, pp. 86–91.

[31] H.-H. Phan, T. T. Nguyen, N. H. Phuc, N. H. Nhan, D. M. Hieu, C. T. Tran, and B. N. Vi, "Key frame and skeleton extraction for deep learning-based human action recognition," in *Proc. RIVF Int. Conf. Comput. Commun. Technol. (RIVF)*, Aug. 2021, pp. 1–6.

[32] Y. Kim and H. Myung, "Gesture recognition with a skeleton-based keyframe selection module," 2021, *arXiv:2112.01736*.

[33] C. Lv, J. Li, and J. Tian, "Key frame extraction for sports training based on improved deep learning," *Sci. Program.*, vol. 2021, pp. 1–8, Sep. 2021.

[34] R. A. Teimoor and A. M. Darwesh, "Node detection and tracking in smart cities based on Internet of Things and machine learning," *UHD J. Sci. Technol.*, vol. 3, no. 1, pp. 30–38, May 2019.
[35] E. Ws, "Center of mass of the human body helps in analysis of balance and movement," *MOJ Appl. Bionics Biomech.*, vol. 2, no. 2, Apr. 2018.
[36] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, 2000.
[37] H. Miyamori and S.-I. Iisaku, "Video annotation for content-based retrieval using human behavior analysis and domain knowledge," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit. (PR)*, Mar. 2000, pp. 320–325.
[38] G. Zhu, C. Xu, Q. Huang, W. Gao, and L. Xing, "Player action recognition in broadcast tennis video with applications to semantic analysis of sports game," in *Proc. 14th ACM Int. Conf. Multimedia*, Oct. 2006, pp. 431–440.

MUHAMMAD ATIF SARWAR received the B.S. and M.S. degrees in computer science from COMSATS University Islamabad, Sahiwal Campus, Pakistan, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree with the EECS International Graduate Program, National Yang Ming Chiao Tung University, Taiwan. His research interests include artificial intelligence, deep learning, and computer vision. His current research to detect activity recognition and actions in a retailers store, sports, and exercise.

YU-CHEN LIN is currently pursuing the master's degree with the National Yang Ming Chiao Tung University, Taiwan. His research interests include artificial intelligence, deep learning, and computer vision.

YOUSEF-AWWAD DARAGHMI received the B.E. degree in electrical and computer engineering from An-Najah National University, in 2002, and the master's and Ph.D. degrees in computer science and engineering from the National Chiao Tung University, Taiwan, in 2007 and 2014, respectively. He is currently an Associate Professor with the Computer Systems Engineering Department, Palestine Technical University–Kadoorie. His research focuses on intelligent transportation systems, vehicular ad hoc networks, and blockchain. He received the Best Paper Award from the International Conference on Intelligent Transportation Systems Telecommunications, in 2012. He served as a Technical Program Committee Member for the International Conference on Connected Vehicles and Expo (ICCVE 2012–2016), the International Conference on Intelligent Transportation Systems Telecommunications (ITST 2012–2018), the International Conference on Signal Processing (ICOSP 2015 and 2016), and the Asia–Pacific Network Operation and Management Symposium (APNOMS 2015.2016). He is a Reviewer of some highly distinguished journals, including IEEE TRANSACTION ON INTELLIGENT TRANSPORTATION SYSTEMS, *IEEE Communication Magazine*, and *IEEE Network Magazine*.

TSÌ-UÍ ÍK (Member, IEEE) received the B.S. degree in mathematics and the M.S. degree in computer science and information engineering from the National Taiwan University, in 1991 and 1993, respectively, and the Ph.D. degree in computer science from the Illinois Institute of Technology, in 2005. He is currently a Professor with the Department of Computer Science and the Director of the Institute of Computer Science and Engineering, National Yang Ming Chiao Tung University. His research focuses on intelligent applications, such as intelligent sports learning and intelligent transportation systems, mobile sensing, machine learning, deep learning, and wireless sensor and ad hoc networks.

He has been a Senior Research Fellow with the Department of Computer Science, City University of Hong Kong. He was bestowed the Outstanding Young Engineer Award by the Chinese Institute of Engineers, in 2009, and the Young Scholar Best Paper Award by IEEE IT/COMSOC Taipei/Tainan Chapter, in 2010. He received the Best Paper Award at ITST 2012. He received a three year Outstanding Young Researcher Grant from the National Science Council, Taiwan, in 2012. In 2020, he received the Sports Science Research and Development Award, MoE, Taiwan. In 2020 and 2021, his research works received the MOST Future Tech Award.

YIH-LANG LI (Member, IEEE) received the B.S. degree in nuclear engineering and the M.S. and Ph.D. degrees in computer science, majoring in designing and implementing a highly parallel cellular automata machine for fault simulation from the National Tsing Hua University, Hsinchu, Taiwan. In 2003, he joined the Faculty of the Department of Computer Science, National Chiao Tung University (NCTU), Hsinchu, where he is currently a Professor. From 1995 to 1996 and from 1998 to 2003, he was a Software Engineer and an Associate Manager with Springsoft Corporation, Hsinchu, where he first completed the development of design rule checking (DRC) tool for the custom-based layout design and then established and led a routing team for developing a block-level shape-based router for the custom-based layout design. His current research interests include physical synthesis, parallel architecture, vehicle navigation, and deep learning. He joined the technical committee of the first CAD contest in Taiwan and served as a committee member for ten years. He has been serving as the Compensation Committee Member and the Independent Director of the Board of Directors for AMICCOM Electronics Corporation, since 2012. He was a recipient of the Japan Society for the Promotion of Science Faculty Invitation Fellowship. He was the Contest Chair of the first CAD Contest at ICCAD, in 2012, and the Technical Program Committee Member of ASPDAC and DAC.