Advances in Science and Technology
Research Journal

Advances in Science and Technology Research Journal, 18(6), 159–176
https://doi.org/10.12913/22998624/191264
ISSN 2299-8624, License CC-BY 4.0

Received: 2024.06.26
Accepted: 2024.08.15
Published: 2024.09.08

Tennis Patterns Recognition Based on a Novel Tennis Dataset – 3DTennisDS

Maria Skublewska-Paszkowska[1*], Pawel Powroznik[1], Edyta Lukasik[1], Jakub Smolka[1]

[1] Department of Computer Science, Lublin University of Technology, ul. Nadbystrzycka 38D, 20-618 Lublin, Poland
* Corresponding author's e-mail: maria.paszkowska@pollub.pl

ABSTRACT

Many scientific studies on tennis stroke recognition are based on datasets, that have been developed for particular studies, using video or motion capture techniques. The importance of such datasets has been increasing due to the athlete performance evaluation needs. The primary contribution of this study is to present a state-of-the-art 3DTennisDS storing four tennis moves: backhand, forehand, volley backhand as well as volley forehand. The strokes were registered using the Vicon optical motion capture and contain a 39 – marker player and a 7 – marker tennis racket models. The potential and quality of this unique dataset has been verified using spatial-temporal graph neural networks, because this type of network topology matches to the human body structure. The presented 3DTennisDS has been compared with two well-known datasets: the THETIS and the Tennis-Mocap. They contain tennis movements in a form of motion capture data, registered using markerless and marker-based systems. The classification of tennis movements has been performed to verify how different types of data acquisition (marker-based and marker-less ones) as well as the structure of the data affect the accuracy of human action recognition. In this study ONI files from THETIS, bvh from Tennis-Mocap and c3d data from 3DTennisDS were considered. Moreover, the impact of input data fuzzification was examined. The obtained results showed that the classification using 3DTennisDS achieved the best results, both for fuzzy and non-fuzzy inputs. These outcomes indicate that the way of capturing data, its preparation and structure have great influence on classification accuracy. The developed 3DTennisDS has a great potential in further motion capture analysis.

Keywords: 3DTennisDS, tennis dataset, tennis strokes, motion capture, graph convolutional networks, fuzzy classification.

INTRODUCTION

The fast progress of computer vision has gained a great impact on interdisciplinary fields, including sports research. Modern, sophisticated equipment enables researchers to obtain more and more accurate data and thus more precise analysis. A high-level understanding of the information gathered in digital images, video or motion capture data allows athletes to maximize their performance, verify their progress, but also to involve people in sports activities with the use of various applications [1]. Modern technologies, such as artificial intelligence, intelligent retrieval, the Internet of Things, as well as machine learning have been widely applied to analyse the way athletes play, compete, and train.

Today technological development provides easier access to up-to-date tools, which make possible the measurement of the three-dimensional movement of tennis players. Various types of motion capture systems allow researchers to perform studies in both: the laboratory settings as well as on the tennis court. Marker-based systems are accurate because retro-reflective or LED markers are attached to the proper parts of the body and indicate their precise positions. Obtained recordings usually need additional post-processing, which is long-lasting. Further, less accurate markerless equipment is more accessible and therefore more often used in the analysis of tennis strokes. Due to the growing popularity of tennis, more and more scientific studies have been conducted. In 2021 almost 87 million people played this sport around the world, which accounted for 1.17% of the world's population [2]. This resulted in increasing scientific works done in the area of plyer movement analysis. Many researchers use existing datasets available online or create their own, for the purposes of conducted studies. Their importance for various sports disciplines has been appreciated [3]. This kind of information is vital to compare the athlete performance using previous and current data. Various types of data storage such as, numerical data, graphics, audio or video recordings, can also be considered as multimedia repositories of sports information. Datasets are exceptionally valuable assets, especially if the information provides timely and efficient data that is adjusted to particular requirements. In addition, the stored data are verified by experts, so they are accurate, which affects the precision of further analysis and numerical experiments.

Based on a literature review, the authors observed the need to collect a tennis stroke data that accurately reflect the movements while performing tennis strokes together with the trajectories and positions of the tennis racket. The following main aims have been defined:

1. Present a state-of-the-art tennis dataset, titled 3DTennisDS, containing the basic tennis moves captured by a Vicon system,
2. Compare the quality of human action recognition based three datasets that represent tennis moves differently: the well-known THETIS dataset based on a Kinect, the Tennis-Mocap dataset gathered bvh data and the 3DTennisDS containing c3d files,
3. Verify how the impact of input data fuzzification influences on classification accuracy.

RELATED WORKS

Studies concerning tennis stroke recognition methods can be found in a number of publications. Analyses have been performed mainly using data obtained from various types of motion capture systems or videos. Usually, the authors use publicly available databases (e.g., THETIS, MADS, and MSR Action 3D) or their own, created for the purposes of the conducted studies.

Action recognition based on motion capture datasets

The martial arts, dancing and sports (MADS) dataset containing both motion capture data and video data for various types of actions, such as arts and sports [4]. Also, tennis movements like serve, forehand and backhand were included. Data were gathered using a seven-camera motion capture system. Each participant had thirty-five markers attached to the body, nineteen of which indicated human body joints. The body shape and pose parameters were evaluated using a personalized depth tracker.

A well-known public THETIS dataset containing tennis movements recorded by a Kinect was presented in [5]. Twelve tennis strokes were recorded: forehand flat, forehand open stands, forehand slice, forehand volley, service flat, backhand, backhand with two hands, backhand slice, backhand volley, service kick, service slice, and smash. Thirty-one amateurs and twenty-four experienced players took part in motion capture sessions. Each stroke was performed several times, which resulted in obtaining 8734 videos in AVI format. The dataset is represented by data such as: depth, RGB, 2D and 3D skeleton, as well as silhouette. It contained skeleton joints as well. Additionally, based on the gathered data, classification of all captured tennis movements was performed.

Many researchers use these twelve tennis moves from this dataset in their studies [6, 7, 8, 9, 10, 11, 12]. One example is tennis movements recognition using the five-layer deep historical long short-term memory (LSTM) network [6]. For indicating the current move the previous frame representing tennis player position were also taken into consideration. The Inception V3 was applied for feature extraction for classification purposes. In [7] the authors created a model for distinguishing these twelve various tennis movements using InceptionResNetV2 and ResNet152V2. For feature extraction based on RGB video frames the CNN-LSTM network was used, while for spatial features from video extraction the Xception was applied.

In [8, 9] all tennis moves from the THETIS dataset were classified using the 3-layered LSTM network. As in previous study, the Inception was used for feature extraction from RGB video frames. Recognition of six tennis strokes using the LSTM classifier was also presented in [10]. The study was performed also on RGB data. The end-to-end network with channel and attention modules was applied for classification. The determining level of expertise of tennis players' actions based on twelve tennis moves from the THETIS dataset was described in [12]. Both shape and motion were considered for k-NN classification with dynamic time warping (DTW) as a distance metric. The same tennis movements were also recognised utilizing linear-chain conditional random fields (CRF) and support vector machine (SVM) [11]. The actions were interpreted as words sequence with direction and speed in region of interest (ROI). Tennis swing and tennis serve were recognized using a convolutional neural network (CNN) in [13]. The network was tested on the medium-sized level MSR Action 3D database. From RGB images the 2D skeleton joints were extracted. Despite the lack of orientation between human limbs, hight recognition accuracy was obtained for all tennis moves, up to 98.10%.

A new approach involving Hilbert embedding-based framework (EHECCO) to extract the nonlinear dependencies for time series classification was presented in [14]. Motion capture (Mocap) data classification covering subject, style and action recognition were performed using three datasets: HDM05, CMU and Tennis-Mocap. Various types of methods were compared for tennis action recognition purposes.

There are a great number of studies where the datasets were created for the purpose of the authors' research. In [15] the dataset of tennis movements for five participants was created using Mocap system. The swinging motions of forehand and backhand strokes were analyzed based on hand positions using a hidden Markov model. The dataset of the forehand swing, slow and fast, performed by novice and intermediate tennis players was described in [16]. The data were captured using a nine-camera eMotion (BTS) SMART-e 900. The movements were performed with a tennis racket without a ball with the frequency set to 50 Hz. Due to various ways of performing forehand strokes, the ROI included from 7 to 13 frames. The classification of basic tennis moves using Learning Vector Quantization was presented in [17]. The dataset of data gathered from the PIQ Robot sports tracker, which used inertial sensor technologies, pressure sensor and a cutting edge microprocessor. The obtained results showed that basic moves, forehand, backhand, serve, and volley, were recognized with an accuracy up of to 90%. However, specific types, such as: flat, slices, and lifted did not achieve the same accuracy, due to the similarity of strokes.

Tennis motion recognition taking into consideration tennis racket orientation may be found in many studies. Tennis moves classification, with deep neural network, including serve, backhand and forehand, together with their various types (e.g. topspin or slice) was described in [18]. The data of 5682 labelled shots were recorded with the wearable SensorTile from 16 amateur players aged 13 to 70. Recognition of six tennis swinging moves using the same type of data utilizing decision tree (DT), SVM, neural networks (NN) as well as k-NN networks was described in [19]. In [20], tennis serves, forehand and backhand were recognised using two classifiers: SVM with the radial basis function kernel and k-NN classifiers. A wireless inertial measurement unit sensor together with an eight-video camera system was used for capturing the data. Another study of forehand, backhand and non-hit classification using the spatial-temporal graph convolutional neural networks (ST-GCN) may be found in [21]. The data were collected using the Vicon Mocap system.

Action recognition based on datasets containing video

Due to the huge amount and general availability of video and streaming data, action recognition is often analyzed in many scientific papers.

Based on video clips from YouTube, a dataset of mimed human actions was created [22]. The Mimetics Dataset gathered 50 human activities. Totally 713 video data were stored. Tennis videos are not divided into separate types of strokes. Action recognition of all human actions was performed using the 3D CNN network as well as different variations of ST-GCN classifiers for 2D and 3D poses. Another commonly used dataset named UTF-101 was applied in [23, 24]. Therein the tennis swing was recognized among other sport activities. This dataset consists of 101 action classes from YouTube. The stored movements were divided into continuous activities and short actions. To the latter tennis swing was assigned. In [25] a markerless motion capture framework for pose estimating was presented. It defined regions of relevant data that were represented as body parts. Two datasets containing tennis games (Tennis-Sense and HumanEva) in the form of videos were used for this purpose. Tennis movements were not divided into specified strokes.

Action detection, utilizing unsupervised learning of semantic events, of tennis movements such as: forehand, backhand, serve, running, obtained from tennis match videos was presented in [26]. For the studies two video sequences of tennis matches were used. The automated detection of winning shots in tennis matches from broadcasted videos was presented in [27]. A new dataset was created containing short duration video sequences of winning and non-winning shots gathered from the final of the 2017 Wimbledon Grand Slam tournament between Roger Federer and Marin Cilic. The detection was performed from a camera mounted behind the opponents. The dataset consisted of 100 examples of sequence divided into two classes “winners” and “no-winners”. 3D Convolutional Neural Networks were applied for the detection purposes.

In [28] an estimation of the tennis players’ results (score and failure) gained in a game was presented. The dataset consisted of 262 singles’ videos. Tennis pose estimation and ball position were taken into consideration for feature detection. The two-class classification by bidirectional LSTM with attention mechanism was applied. An unsupervised procedure, with frame level annotation, to recognize action phases in tennis videos was described in [29]. For the studies 314 tennis-points videos obtained from video-commentary dataset of London Olympics 2012 were used [30]. The analysis began from the serve and ended with winning a point.

The way of tennis stroke prediction during matches utilizing player’s pose as well as position was presented in [31]. The study involved automatically labelling. The dataset containing videos of professional tennis matches was used. LSTM was used for extracting features (player’s position and court’s lines), while recurrent neural network (RNN) was applied for prediction of stroke directions. Recognition and classification using Mahalanobis distance of forehand, backhand and serve were presented in [32]. The data were gathered from three professional right-handed tennis players by a single AXIS 215 PTZ camera. Sixty serves, sixty forehands and forty backhands were recorded. The pose estimation using Convolutional Neural Network for three tennis movements was presented in [33]. For the purpose of this study a new dataset of 310 images representing tennis athletes together with body parts annotations was created. The TenniSet created from five matches at the 2012 London Olympics, was obtained from YouTube was presented in [34]. A sequence of events corresponding to the match in a way of attributes were added to the dataset. For the frame classification the VGG16 network was used, while for the purpose of events and actions recognition both a bidirectional and a forward direction RNNs were applied. A dataset created from the broadcast tennis video at the 2017 Summer Universiade was presented in [35]. Totally, 81 game-related clips were gathered. Each clip started with serve and finished with gaining a point. The tennis ball was detected and its move was tracked. Three trajectory patterns were defined: flying, hit, and bouncing. The 13 layers of VGG-16 were applied for ball trajectory pattern classification, while the 14–24 layers of DeconvNet were used for ball trajectory prediction. The dataset was created from TV broadcasts for five sports disciplines. Volleyball, basketball, tennis, cricket and football were used for classification sport disciplines using SVM [36]. From the video the features were extracted in the way of edge information from sub-bands.

The performing tennis serves was considered in [37]. It was said that players eagerly took risks while performing their first serve. The results showed that for this type of tennis move performed by tennis pro there is a strategy how to do first or second serve minimizing the risk. The study was performed on the dataset containing the international skilful tennis games from 2005 to 2009, gathered from the Grand Slam and Association of Tennis Players tournaments, or lower-level International Tennis Federation tournaments. In total, 3188 games corresponding to 69.1360 serves were taken into consideration. The tool for planning matches against an opponent was presented in [38]. Each ball trajectory, gathered from Hawk-Eye system, of various types of serves was represented as two sub-trajectories, from which the features were extracted such as: angle, location and impact speed. Tracking a tennis ball during the match using a machine learning approach was presented in [39]. This method utilized random forest segmentation for identifying the ball. The algorithm was verified on a dataset created from Roger Federer play from Youtube.

The identification of racket sport (including tennis and badminton) tactics, its analysis, as well as the player’s progression was presented in [40]. The described algorithm for multidimensional patterns was performed on badminton and tennis datasets.

The Kernelised Linear Discriminant Analysis with annotation transfer learning was applied for hit, serve and non-hit identifications [41]. The basic tennis groundstrokes classification utilizing SVM from a video may be found in many studies [42, 43, 44].

MOTION CAPTURE DATASETS

For human action recognition focused on tennis movements the following motion capture datasets were taken into consideration: THETIS, Tennis-Mocap and final, 3DTennisDS, created for the purpose of this experiments. All of the above-mentioned datasets are publicly available. They contain tennis stroke motion capture trajectories, recorded in different ways, and saved in various data formats. The THETIS is one of the best-known datasets captured using markerless system, while the Tennis-Mocap was chosen as an example of marker-based one.

THETIS

The THETIS dataset was presented in 2013 [5]. Twelve various tennis movements were recorded using Microsoft's Kinect. Tennis strokes were carried out by thirty-one amateurs and twenty-four professional athletes. The dataset includes: backhand (one-handed, two-handed, slice), volley (backhand and forehand), forehand (flat, open stands, slice), service (flat, kick, slice) and smash. The following data formats are available: ONI files, depth and RGB videos, silhouette, skeleton 2D and 3D videos, and skeleton joints. It is one of the most cited datasets and constitutes the background for many studies. The ONI files present the model of the tennis player consisting of 15 points (Figure 1).

Tennis-Mocap

For the purpose of action classification and movement analysis the Tennis-Mocap dataset was created in 2020 [14]. 17 athletes of the Caldas-Colombia tennis league were the participants included in the research. Optitrack Flex V100 including six cameras with the frequency set to 100 Hz was used for tennis movement capturing. Thirty-four markers were attached for collecting information about body joints. All participants were asked to hit the ball with constant velocity. Their movements were supposed to be as much similar during the match as it was possible. They performed series of the following strokes: groundstrokes (forehand and backhand), serve, volleys and smash.

3DTennisDS

In this part, we present the process of creating a unique 3DTennisDS dataset containing recordings of basic tennis strokes acquired by Vicon, an optical motion capture system. The whole methodology (see Figure 2) consists of capturing

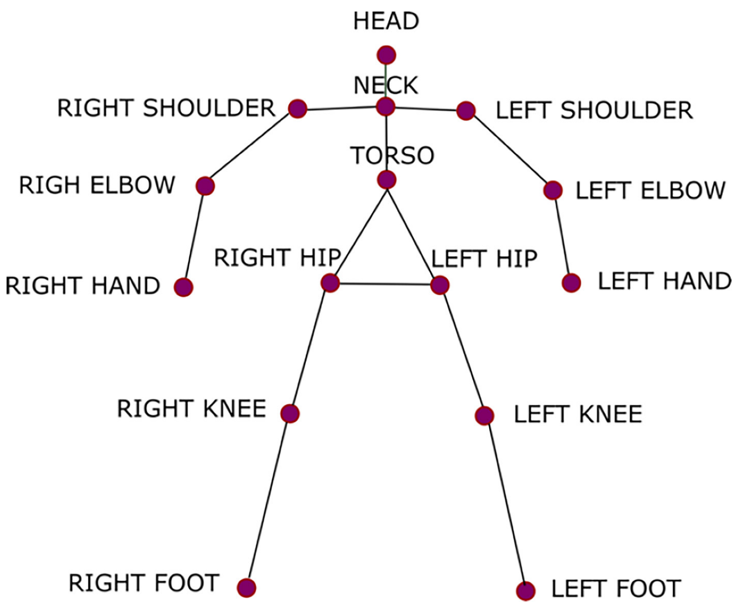

Figure 1. THETIS subject model

tennis moves, data post-processing and data preparation for public availability.

Capturing tennis movements

The acquisition of the basic tennis strokes was carried out in the motion capture laboratory at the Lublin University of Technology. The basic tennis movements, forehand, backhand and volley, were registered utilizing the optical 8 – camera Vicon Mocap system with the frequency set to 100 Hz. Both the tennis player and the tennis racket were captured.

Ten tennis professionals were the participants in this study. In total, 39 retroreflective markers were attached to their bodies using hypoallergenic double-sided tape, according to the Plug-in Gait Model (PiG) [45]. Then, they were measured for the purpose of creating and scaling new subjects in the Vicon Nexus software. The parameters were collected for left and right parts of body. Additionally, seven retroreflective markers were also fixed to the tennis racket (Figure 3). That resulted in reconstructing the shape of the racket and obtaining the racket's trajectory.

Participants carried out four tennis movements: backhand, forehand, volley (forehand and backhand). Forehand and backhand were performed while the tennis player was moving. The participant tried to avoid a small bollard placed on the floor. This resulted in obtaining more natural moves, similar to those performed directly on the court. At first, ten forehand strokes without a ball were carried out, followed by ten backhand strokes without a ball. Secondly, these activities were done with a ball. Volleys were performed in front of a tennis net. The participant was asked to perform these types of moves alternately.

Data post-processing

Each recording was reviewed and all irregularities were corrected using the Vicon Nexus software. The post-processing consisted of four

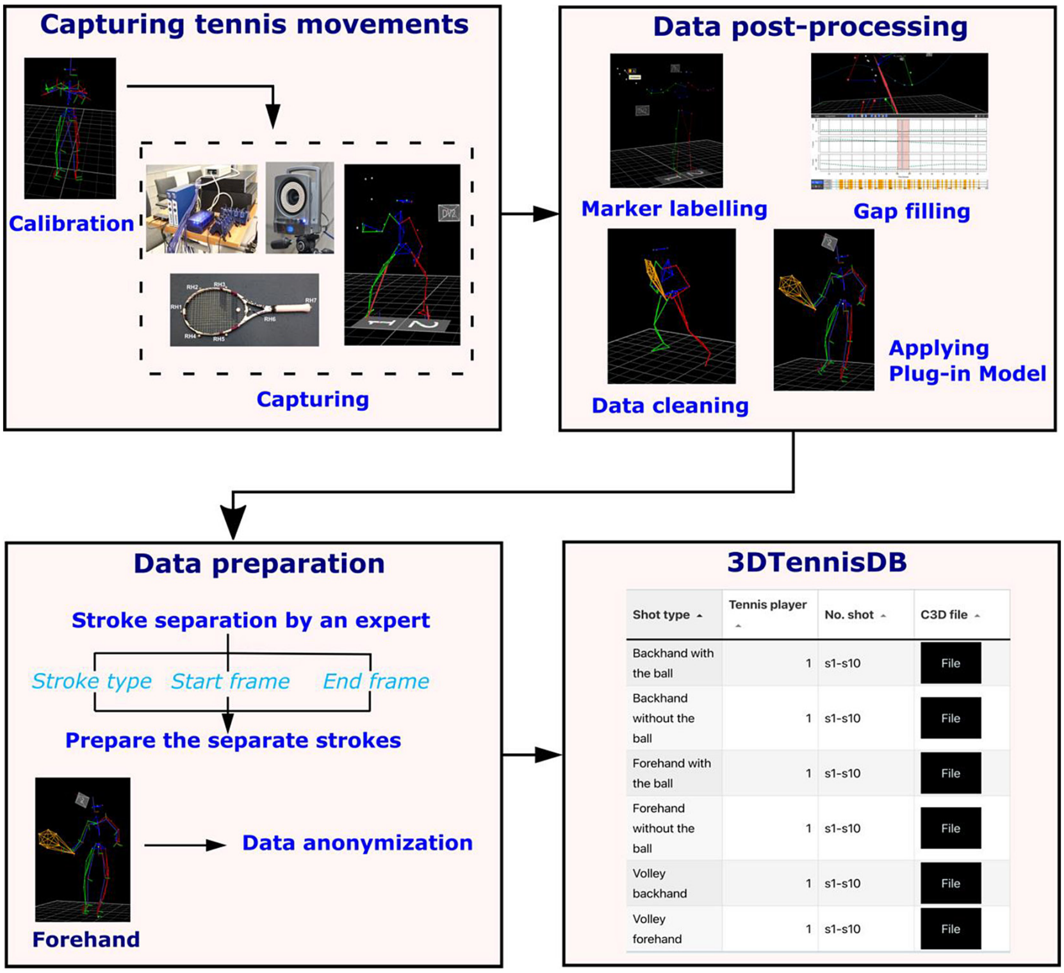

Figure 2. Schema for creating 3DTennisDS

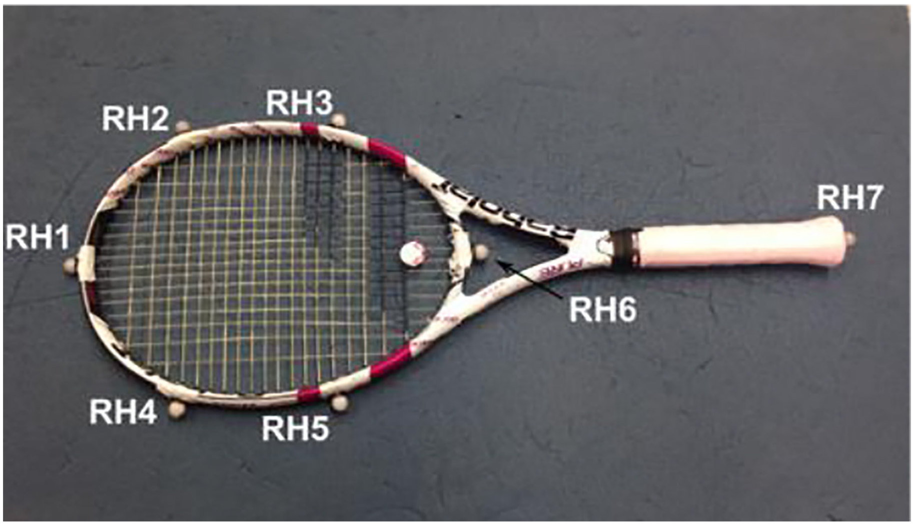

Figure 3. Placement of markers on a tennis racket

steps (Figure 2): marker labeling, gap filling using interpolation methods, data cleaning (e.g., deleting all unlabeled markers), and applying the PiG model (only for the participant subject). A new subject was developed for the racket. The corrected recordings have been exported as c3d files. First, all markers in whole frames were labeled by PiG model and the model of the tennis racket. Second, missing markers were interpolated based on the marker's positions from previous and following frames. Third, all unlabeled markers were deleted - all additional markers that appeared as results of reflections or other anomalies were removed. Fourth, the PiG model outputs were obtained. The final c3d file consisted of the 3D marker's positions as well as joints, angles and moments.

Data preparation

The post-processed c3d files contained several tennis strokes. A tennis expert separated consecutive tennis moves. The start and end frames of each move were determined based on both the player's body and the tennis racket positions. As a result, the whole recording was divided into smaller ones, dedicated to one specific tennis move. Before making the data public, it was still necessary to make it anonymous. By removing sensitive data.

3DTennisDS access

The 3DTennisDS dataset is available for public use at https://tennisdb.cs.pollub.pl/. A brief description of the dataset and its terms of use have been prepared. The whole dataset was published in the form of a table (see Figure 2) divided by the tennis participants and the type of moves. For each player the number of strokes is specified. Additionally, the link to the Plug-in Gait model was made available. The users may also download the racket's model.

GRAPH CONVOLUTIONAL NETWORKS

In this study, the tennis strokes recognition was performed with the ST-GCN, which architecture is depicted in Figure 4.

Spatial temporal graph

Three-dimensional coordinates for the selected parts of the player's body for the successive frames were generated from the c3d tennis strokes files. In order to compare three-dimensional data from three various databases, a very important task was to ensure their consistency. To obtain corresponding markers, new positions from the 3DTennisDS database were computed using interpolation. For example, from four markers attached to the head, one point was calculated. A similar situation took place for the determination of the hip positions on the basis of two markers from this model. The interpolation process was also applied for the Tennis-Mocap dataset. Based on the right and left collars a new point was indicated. The obtained points were used to build the graph containing nodes and the connections among them as joints. Considering the sequence of these positions changing in time, a spatial time

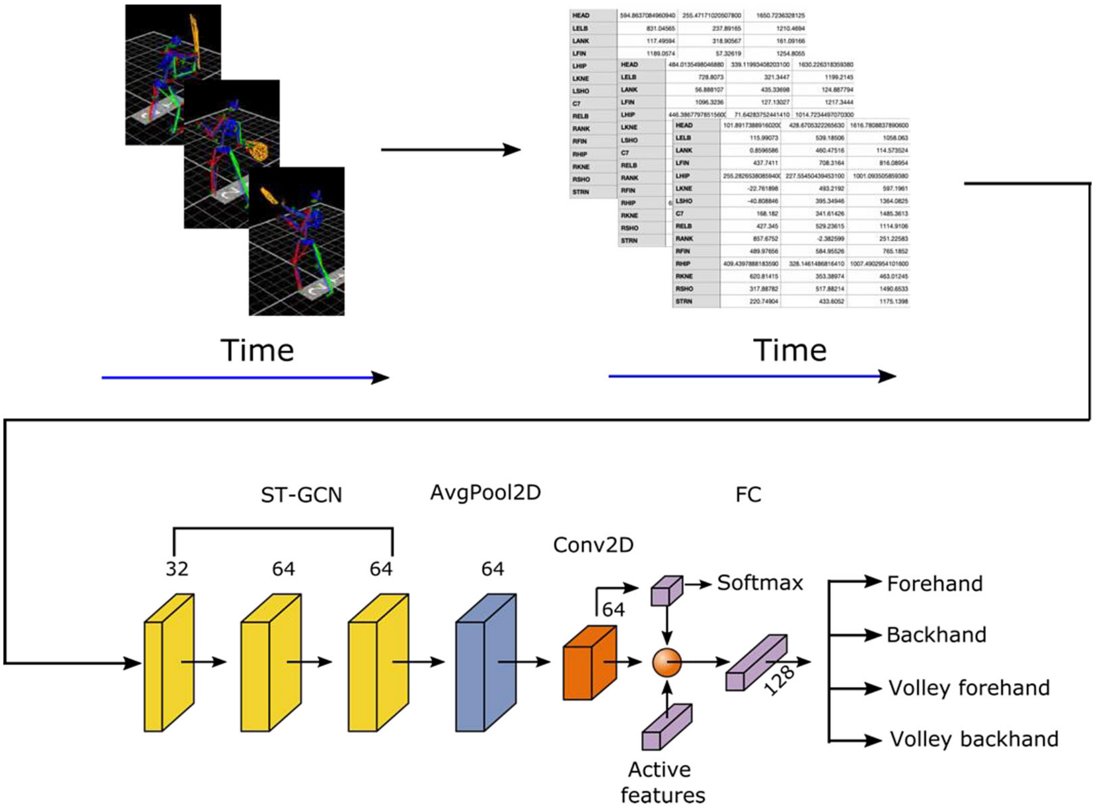

Figure 4. Scheme of applied ST-GCN classifier

graph corresponding to the time representation was built. Thus, the structure of the input data is expressed in the form of a graph G = (V, E) and consists of N nodes corresponding to N joints and changes in their position in time. The node set describes joints in a skeleton (Equation 1).

$$V = v_{ti} |\ t = 1, \cdots, T, i = 1,, \cdots, N \tag{1}$$

Tennis strokes recognition

The vector consisted of graph nodes is given as the input to the neural network. Base on it the operation of layer multiplication of the proposed spatial and temporal graph is performed. As a result, the features are extracted and a fixed-size vector is obtained. The Softmax classifier is utilized in the final stage of the tennis movement classification. The stochastic gradient descent method was applied for network learning.

The ST-GCN solution is constructed utilizing convolutional neural networks. Due to the basic convolutional operation a set of three-dimensional features is obtained from the network input. By specifying proper data padding the input as well as output data are the same size. That is why the convolution operation may be defined as [46]:

$$f_{out} = \Sigma_K \Sigma_K f_{in} p(x) * \omega \tag{2}$$

where: p – the sampling function and ω – the weights function.

It should be stated that the weight function is irrelevant to the location of the x point. A standard convolution is therefore achieved by encoding the rectangular grid in $p(x)$. Detailed explanations can be found in [47]. The sampling function can be defined on the neighbour set $G(v_{ti}) = \{v_{tj} \mid d(v_{tj}, v_{ti}) \leq D\}$ of node v_{ti}, where $d(v_{tj}, v_{ti})$ indicates the minimum length of any path from v_{tj} to v_{ti}. In this research D equals 1. Considering the above it can be written that [46]:

$$p : G(v_{ti}) \longrightarrow V \tag{3}$$

That is why, for graphs, dependence Equation 2 can be described as follows [46]:

$$f_{out} = \Sigma_{v_{tj} \in G(v_{ti})} f_{in}(v_{tj}) * \varpi(v_{tj}) \tag{4}$$

In this study, the ST-GCN method, presented by [48], was applied for tennis movements recognition. In each frame of Mocap data, the connections between joints (presented as matrix I) were defined utilizing Adjacency matrix (A). This relation, for a single frame, can be defined as Equation 5 [46]:

$$f_{out} = \Lambda^{-\frac{1}{2}}(A+I)\Lambda^{-\frac{1}{2}}f_{in}W \tag{5}$$

where: $\Lambda^{ii} = \sum_j (A^{ij} + I^{ij})$ and W is the weight vector.

In this study, the input feature map is regarded as a tensor of *(C,V,T)* dimensions. The graph convolution is performed as a two-dimensional convolution matrix, multiples with normalized, two-dimensional adjacency matrix.

The classifier proposed for tennis movement recognition is created with the following layers: a three-layer ST-GCN (consisting of 32, 64 and 64 kernels, respectively), pooling, convolutional layer (Figure 4). After the third layer, the average pooled data of joints and temporal directions are forwarded to a 1 × 1 convolutional layer. The final is the four-dimensional layer, followed by the Softmax function, corresponds to the four recognized tennis movements.

Fuzzy approach

Due to the lack of sharp boundaries between strokes, it was decided to fuzzify the input.

Definition 1. A non-empty fuzzy set f_s can be understood as an ordered pair (f_s, η_{fs}) where n_{fs} is a membership function $n_{fs} : f_s \to [0, 1]$, which allows to perform the fuzzyfication operation. η_{fs} assigns to each element x in f_s a degree of membership, $0 \leq \sigma \leq 1$ [49].

Definition 2. A fuzzy relation on f_s is a fuzzy subset of $f_s x f_s$. A fuzzy relation η_{fs} on f_s is a fuzzy relation on the fuzzy subset σ, if $\eta_{fs}(x, y) \leq \sigma(x) \wedge \sigma(y)$ for all x, y from f_s and $\wedge$ stands for minimum. A fuzzy relation η_{fs} on f_s is said to be symmetric if $\eta_{fs}(x, y) = \eta_{fs}(y, x)$ for all $x, y \in f_s$ [49].

Definition 3. A fuzzy graph is a pair G: (σ, η_{fs}) where σ is a fuzzy subset of f_s, η_{fs} is a symmetric fuzzy relation on σ [49].

Definition 4. (σ', η'_{fs}) is a fuzzy subgraph of: (σ, η_{fs}) if $\sigma' \subseteq \sigma$ and η'_{fs} [49].

Tennis movements may be recognized with "uncertain" phenomena [50]. Adding fuzzification process to graph neural networks may be taken as a pattern how people perceive these moves. This process blurs the sharp boundaries between the investigated sets. Applying both fuzzy rules and the reverse operation allows one to change one state to another. The fuzzification of the input data was performed utilizing the membership functions:

$$\eta_{Rf_s}(x) = \begin{cases} 0, (x > d) \\ \frac{d-x}{d-c}, c \leq x \leq d \\ 1, x < c \end{cases} \tag{6}$$

$$\eta_{f_s}(x) = \begin{cases} 0, (x < a) || (x > d) \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, b < x < c \\ \frac{d-x}{d-c}, c \leq x \leq d \end{cases} \tag{7}$$

$$\eta_{Lf_s}(x) = \begin{cases} 0, (x < a) \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, x > b \end{cases} \tag{8}$$

where: a, b, c, d denote trapezoidal function parameters and $a<b<c< d$. In the case of Equation 6 a=b=–∞ and in Equation 8 c=d=∞.

EXPERIMENTS AND RESULTS

Used measures

In this section, the potential of the created 3DTennisDS dataset was verified by applying the ST-GCN classifier. The three-dimensional coordinates of the markers attached to the models of a tennis player and a tennis racket were taken from the c3d files. The described dataset was compared with the well-known THETIS dataset and the Tennis-Mocap in order to verify how various types of data acquisition affect the accuracy of the human action recognition. In this study all ONI files from THETIS, all bvh files from Tennis-Mocap, as well as all c3d data from 3DTennisDS consisting forehand, backhand, volleys (forehand and backhand) were involved in the study. Moreover, the impact of fuzzification input data for both datasets was examined. The experiment consisted of the following steps: 1) Define the classes for four types of tennis movements. 2) Adapt Mocap data to a uniform model allowing comparison of available databases (consonant with the defined classes). 3) Divide dataset to three subsets: training, testing and validation in 60%, 20%, 20% proportion, respectively. 4) For applying fuzzification, the input data were prepared using trapezoidal functions. 5) The 3DTennisDS dataset was published at https://tennisdb.cs.pollub.pl/. 6) Network learning. 7) The network was tested over twenty trials. 8) A confusion matrix was created and were computed [50].

Evaluation of the tennis dataset

The experiments described in this study concerned tennis movement recognition using three various datasets containing tennis motion strokes. Four main tennis moves were taken into consideration. The first dataset, the THETIS, gathered movements captured using a Microsoft Kinect, using markerless method. The other two sets contained data recorded with the marker-based method at the same frequency, which was set to 100 Hz. However, they differed in marker placement model as well as their representations. The second datatset, the Tennis-Mocap, stored data in the form of bvh file. Due to the growing demand for databases containing accurate data, the authors had developed a new set of tennis strokes, the 3DTennisDS, registered with the Vicon motion capture system using 39 markers attached according to the PiG model. The data were stored as.c3d files in the third dataset. Since each of the datasets contained data represented in the form of various number of points (markers), they were adapted to the THETIS dataset to obtain fair comparison. From each file the three-dimensional coordinates of the corresponding markers were taken. In the Tennis-Mocap dataset left and right collar were interpolated into one point. In the 3DTennisDS LPSI and LASI markers, as well as RPSI and RASI were interpolated into left and right hips, respectively. Four markers placed on the head were also interpolated into one. Additionally, from c3d data only corresponding markers' coefficients to the THETIS model were taken into consideration. Additionally, in this study the tennis racket, consisting of seven markers, was included.

In the study the authors investigated how the input data fuzzification affects the quality of tennis moves classification. In Table 1 the obtained accuracy of the tennis moves from three datasets is presented. The tennis stroke recognition for the dataset created using the Vicon Mocap system has reached a higher average accuracy than the dataset defined using the Microsoft Kinect and the Optitrack Flex systems. The obtained maximum and minimum accuracy values also favor of the 3DTennisDS dataset.

It should be stated that data from marker-based systems allow for higher accuracy than in the case of a markerless system. The obtained accuracy for the 3DTennisDS is higher than for the THETIS dataset (7.45% for non-fuzzy input and 7.20% for fuzzy input, respectively). In case of the Tennis-Mocap dataset the achieved accuracy is also higher for the 3DTennisDS (4.59% for non-fuzzy input and 5.93% for fuzzy input, respectively). The accuracy results obtained on two datasets (THETIS and Tennis-Mocap) are comparable. They small difference speaks in favor of the Tennis-Mocap dataset (2.86% for non-fuzzy input and 1.27% for fuzzy input, respectively). The values of the standard deviation for three datasets means lower variability of the distribution, and also that the obtained results do not differ significantly from each other.

Analyzing the results obtained in Table 2 for individual tennis strokes, it should be noted that the highest accuracy was obtained for the newly created 3DTennisDS for all tennis moves. Higher values were obtained for fuzzy data. This is due to no distinct boundaries between strokes (e.g., forehand and volley forehand). These strokes were classified with higher accuracy. It should be noted that for all analyzed datasets, all strokes were classified at a similar level within the particular dataset. In Table 3 the precision for all analyzed tennis strokes is gathered. This parameter stands for the ratio of correctly classified elements (TP) to all elements marked by the used classifier as ($TP + FP$). The mean precision is higher for all strokes after applying fuzzification to the input for all datasets. A better precision was achieved for the 3DTennisDS dataset in all

Table 1. Obtained accuracy results

Dataset	Type of input	Mean	Max	Min	±SD
THETIS	Non-Fuzzy	74.40%	80.00%	69.00%	3.18%
	Fuzzy	80.40%	84.00%	77.00%	2.18%
3DTennisDS	Non-Fuzzy	81.85%	85.00%	77.00%	2.59%
	Fuzzy	87.60%	92.00%	82.00%	3.07%
Tennis-Mocap	Non-Fuzzy	77.26%	81.90%	73.00%	3.06%
	Fuzzy	81.67%	84.60%	79.10%	2.78%

Table 2. Obtained accuracy results for individual strokes

Dataset	Type of input	Stroke	Mean	Max	Min	±SD
THETIS	Non-fuzzy	Forehand	73.21%	80.00%	68.39%	3.26%
		Backhand	73.47%	79.37%	68.02%	3.49%
		Volley forehand	75.47%	79.85%	70.14%	3.51%
		Volley backhand	74.08%	80.00%	68.46%	3.57%
	Fuzzy	Forehand	80.68%	84.28%	76.52%	2.83%
		Backhand	80.81%	84.76%	76.52%	2.83%
		Volley forehand	80.10%	83.49%	76.66%	2.07%
		Volley backhand	81.43%	84.38%	76.45%	2.23%
3DTennisDS	Non-fuzzy	Forehand	80.91%	85.93%	76.43%	2.78%
		Backhand	80.39%	85.81%	77.05%	3.20%
		Volley forehand	81.58%	85.94%	76.44%	2.93%
		Volley backhand	81.09%	85.22%	76.17%	2.98%
	Fuzzy	Forehand	87.07%	91.93%	81.31%	3.16%
		Backhand	87.78%	92.04%	82.04%	3.13%
		Volley forehand	87.97%	92.48%	82.22%	3.04%
		Volley backhand	85.42%	92.33%	81.33%	3.05%
Tennis-Mocap	Non-fuzzy	Forehand	76.98%	85.50%	73.05%	2.47%
		Backhand	77.10%	81.61%	73.23%	3.01%
		Volley forehand	78.13%	81.98%	73.30%	3.10%
		Volley backhand	77.79%	81.55%	73.11%	2.47%
	Fuzzy	Forehand	84.62%	91.93%	79.14%	2.92%
		Backhand	82.47%	84.84%	79.15%	3.93%
		Volley forehand	81.20%	84.30%	79.25%	3.34%
		Volley backhand	82.25%	84.82%	79.54%	3.52%

cases than for the THETIS and the Tennis-Mocap ones. In Table 4 the recall stands for the ratio of correctly recognized elements from class (*TN*) to all elements from that class. All tennis strokes gained better mean recall for 3DTennisDS dataset both for fuzzy and non-fuzzy input except non-fuzzy backhand in Tennis-Mocap dataset. In Table 5 the F1 score, a harmonic mean, of the precision and recall measures is presented. For all tennis strokes for both fuzzy and non-fuzzy input, the 3DTennisDS dataset achieved better results. The highest difference between 3DTennisDS and THETIS datasets reached 8.86% for fuzzy forehand, while the lowest difference achieved 6.55% for non-fuzzy backhand. While for the 3DTennisDS and the Tennis-Mocap datasets the highest difference was gained 8.32% for fuzzy volley forehand and the lowest one, 3.29%, for non-fuzzy backhand.

The obtained classification results have shown that in all cases the best results are achieved for the 3DTennisDS. Only for this dataset the maximum F1 result was higher than 90% for all fuzzy strokes. The confusion matrices for the 3DTennisDS, Tennis-Mocap and the THETIS datasets, with applying input data fuzzification process and without it are presented in Fig. 5–7. As it can be seen, the fuzzification of the input data increases the classification results: for THETIS 2.21–2.98%, for Tennis-Mocap 4.42–6.47%, while for 3DTennisDS 5.23–7.44%. In case of fuzzy approach, all strokes from 3DTennisDS achieved more effective stroke recognition, up to 11.47%, in comparison to the THETIS dataset. In case of a non-fuzzy input, all strokes were better classified from the 3DTennisDS than the THETIS with the difference up to 7.65%. It can be clearly noticed that for all datasets, the following strokes are misclassified: forehand with volley forehand and backhand with volley backhand. On the other hand, the smallest percentage of misclassifications was obtained in case of the 3DTennisDS.

Analyzing the obtained results, it can be stated that accurate methods of data acquisition, as in the case of an optical motion capture system,

Table 3. Obtained precision results for individual strokes

Dataset	Type of input	Stroke	Mean	Max	Min	±SD
THETIS	Non-fuzzy	Forehand	79.30%	84.21%	74.19%	2.75%
		Backhand	77.64%	82.47%	72.63%	2.70%
		Volley forehand	73.34%	79.21%	68.32%	3.28%
		Volley backhand	72.60%	78.34%	67.00%	3.40%
	Fuzzy	Forehand	84.41%	88.54%	81.05%	1.99%
		Backhand	82.93%	86.60%	79.59%	2.18%
		Volley forehand	79.44%	83.17%	76.24%	2.06%
		Volley backhand	79.03%	84.85%	75.49%	2.59%
3DTennisDS	Non-fuzzy	Forehand	85.60%	88.54%	81.05%	2.33%
		Backhand	84.46%	87.63%	80.21%	2.53%
		Volley forehand	80.96%	84.16%	73.24%	2.57%
		Volley backhand	81.17%	85.86%	75.49%	3.60%
	Fuzzy	Forehand	92.91%	97.87%	85.42%	5.07%
		Backhand	89.60%	93.88%	84.54%	2.75%
		Volley forehand	87.22%	92.00%	80.39%	3.46%
		Volley backhand	87.62%	92.00%	80.40%	4.25%
Tennis-Mocap	Non-fuzzy	Forehand	81.81%	85.91%	77.86%	2.72%
		Backhand	77.40%	81.88%	73.51%	2.94%
		Volley forehand	75.49%	80.67%	71.05%	3.20%
		Volley backhand	78.49%	82.88%	73.97%	2.94%
	Fuzzy	Forehand	85.61%	88.11%	83.69%	3.47%
		Backhand	82.18%	85.14%	79.73%	3.75%
		Volley forehand	80.29%	83.44%	77.48%	2.79%
		Volley backhand	82.61%	85.61%	80.13%	2.81%

Table 4. Obtained recall results for individual strokes

Dataset	Type of input	Stroke	Mean	Max	Min	±SD
THETIS	Non-fuzzy	Forehand	70.55%	76.92%	65.09%	3.51%
		Backhand	76.10%	80.81%	71.88%	2.75%
		Volley forehand	79.81%	84.21%	74.19%	2.87%
		Volley backhand	76.62%	82.47%	71.13%	3.10%
	Fuzzy	Forehand	77.29%	84.16%	72.90%	3.04%
		Backhand	81.45%	84.00%	78.57%	1.88%
		Volley forehand	84.45%	88.42%	81.25%	1.83%
		Volley backhand	82.86%	86.60%	79.38%	2.16%
3DTennisDS	Non-fuzzy	Forehand	79.85%	85.00%	72.64%	4.01%
		Backhand	82.38%	85.00%	78.58%	2.20%
		Volley forehand	85.83%	89.47%	81.91%	2.26%
		Volley backhand	84.21%	87.50%	79.38%	2.58%
	Fuzzy	Forehand	86.45%	91.09%	78.85%	4.19%
		Backhand	88.27%	92.93%	82.83%	3.37%
		Volley forehand	93.09%	98.90%	85.42%	4/91%
		Volley backhand	89.56%	93.88%	83.67%	2.88%
Tennis-Mocap	Non-fuzzy	Forehand	74.70%	79.74%	70.78%	3.23%
		Backhand	83.04%	86.52%	79.56%	2.23%
		Volley forehand	79.21%	84.02%	75.00%	3.12%
		Volley backhand	76.85%	81.21%	72.08%	3.20%
	Fuzzy	Forehand	79.705	82.89%	76.62%	2.91%
		Backhand	86.57%	88.73%	84.29%	1.29%
		Volley forehand	83.21%	86.90%	80.82%	2.74%
		Volley backhand	81.36%	84.50%	78.52%	2.87%

Table 5. Obtained F1 score results for individual strokes

Dataset	Type of input	Stroke	Mean	Max	Min	±SD
THETIS	non-fuzzy	Forehand	74.66%	80.40%	69.35%	3.17%
		Backhand	76.88%	81.63%	72.25%	2.68%
		Volley forehand	76.44%	81.63%	71.13%	3.02%
		Volley backhand	74.55%	80.40%	69.00%	3.24%
	fuzzy	Forehand	80.69%	86.29%	77.00%	2.54%
		Backhand	82.18%	85.28%	79.38%	1.99%
		Volley forehand	76.44%	81.63%	71.13%	3.02%
		Volley backhand	74.55%	80.40%	69.00%	3.24%
3DTennisDS	non-fuzzy	Forehand	82.61%	86.73%	76.62%	3.22%
		Backhand	83.41%	86.29%	79.38%	2.35%
		Volley forehand	83.32%	86.73%	78.97%	2.41%
		Volley backhand	82.65%	86.29%	77.39%	3.08%
	fuzzy	Forehand	89.55%	94.36%	82.00%	4.50%
		Backhand	88.93%	93.40%	83.67%	3.05%
		Volley forehand	90.05%	94.84%	82.83%	4.08%
		Volley backhand	88.57%	92.93%	82.00%	3.57%
Tennis-Mocap	non-fuzzy	Forehand	78.09%	82.71%	74.15%	3.00%
		Backhand	80.12%	84.14%	76.76%	3.56%
		Volley forehand	77.48%	82.31%	72.79%	3.16%
		Volley backhand	77.66%	82.03%	73.22%	4.06%
	fuzzy	Forehand	82.54%	85.42%	80.00%	2.72%
		Backhand	84.31%	86.90%	81.94%	2.48%
		Volley forehand	81.73%	85.14%	79.32%	3.75%
		Volley backhand	81.98%	85.03%	79.32%	3.82%

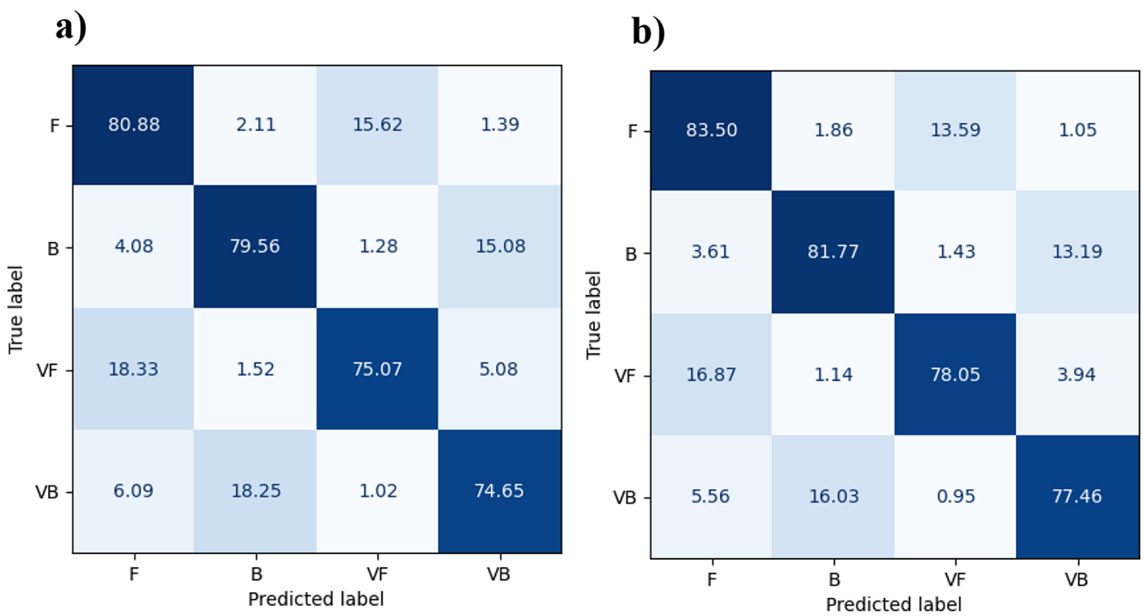

Figure 5. Confusion matrices for THETIS dataset (a) non-fuzzy input (b) with fuzzy input

influences the classification accuracy. Moreover, the applied fuzzification for the more accurate motion capture data significantly increases the obtained results. The correctness of proposed classifier was verified on a basis of Leave-One-Out Cross-Validation (LOOCV). This method allows one to deliver unequivocal information about developed model. However, this is a

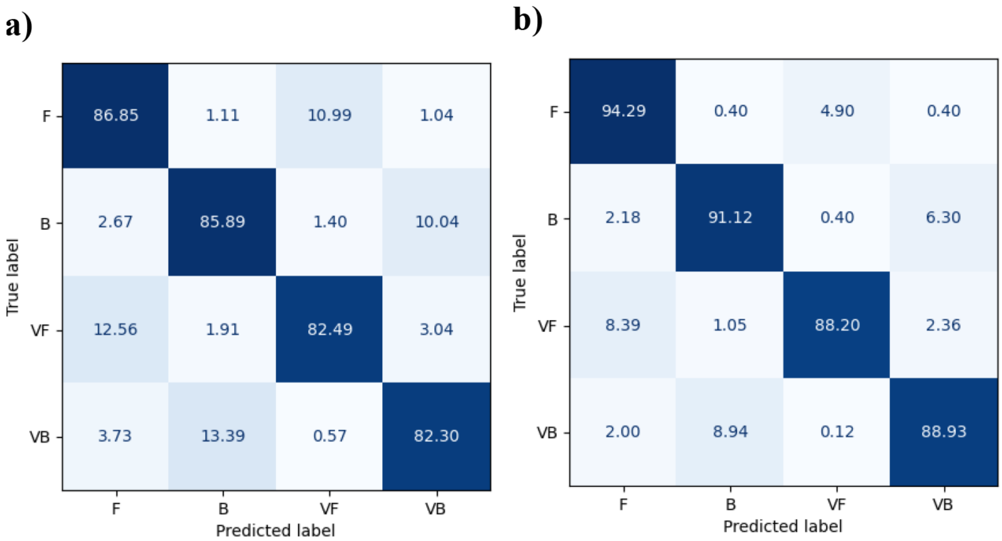

Figure 6. Confusion matrices for 3DTennisDS dataset (a) non-fuzzy input (b) with fuzzy input

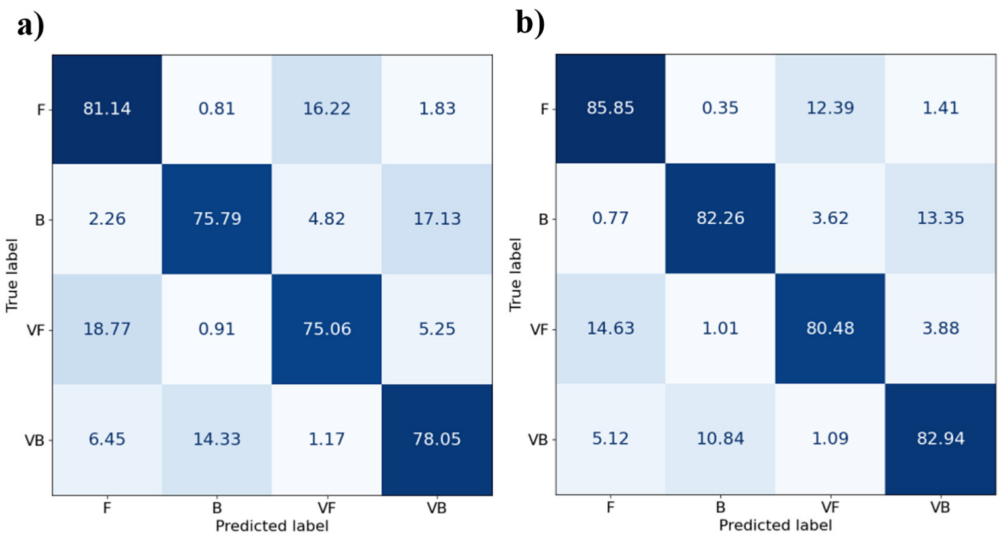

Figure 7. Confusion matrices for Tennis-Mocap dataset (a) non-fuzzy input (b) with fuzzy input

computationally complex and time-consuming approach involving the root mean squared error (RMSE) for *n* tests:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

where: n is the number of tests, y_i is true value, $\hat{y}_i$ denotes predicted value.

The LOOCV values were obtained for the THETIS dataset as: 9.05% ± 6.38% and 10.96% ± 5.48% for non-fuzzy input and input with fuzzification, respectively. The 3DTennisDS obtained 8.99% ± 5.81% and 7.74% ± 4.11% for non-fuzzy input and input with fuzzification, respectively. The Tennis-Mocap achieved 9.46% ± 5.56% and 8.51% ± 4.46%, respectively.

Comparison with the state-of-the-art

A great number of studies about tennis movements recognition and classification performed on signal, sensor, video, images and three-dimensional data may be found in the literature. Many types of movements were analyzed, such as: forehand, backhand, serve, volley, smash, no-hit, as well as various ways of performing them. Analyzing the state-of-the-art (Table 6), it is obvious that the classification performance depends on the type of data. As far as the authors

Table 6. Results comparison with the state-of-the-art

Data/Dataset	Type of input	Classified types of tennis move	Methos	Accuracy	Paper
SensorTile	Signal	F, B, S	DNN	94–97%	[18]
		F, B	SVM	90.82–98.86%	[19]
			NN	98.76–100%	
			DT	84.69–95.54%	
			RF	93.75–98.96%	
			k-NN	87.76–99.44%	
IMU	Sensor	F, B, BS, S, SM	Pan Tompkins algorithm	80.60–98.10%	[52]
THETIS	Video	B, V, F, S, SM	LSTM	81.23–89.42%	[11]
			SVM	51.20%	
			CRF	86..44%	
THETIS	Video	B, V, F	Deep Historical LSTM	62%	[6]
HMDB51		S, SM		54%	
THETIS	Video	B	LSTM	70.17–97.67%	[10]
KTH		V, S, SM			
THETIS	Video	B, F, V	SVM	53.08–60.23%	[5]
KTH		S, SM		90.65%	
KTH	Video	S, H, NH	KLDA	73.34–92.29%	[41]
Broadcast	Video	F, B	SVM	90.21%	[42]
				87.10%	[43][44]
Mixed	Signal, video	F, B, S	SVM	89.69–97.02%	[20]
				82.43–88.36%	
			k-NN	89.41–93.44%	
				84.73–100%	
Vicon	Image	F, B, NH	ST-GCN	64.10–74.30%	[21]
Vicon with fuzzy input	Image	F, B, NH	ST-GCN	86.30–87.30%	[21]
			A3T-GCN	86.90–93.82%	[53]

Note: F–forehand, B–backhand, S–serve, BS–backspin, SM–smash, V–volley, H–hit, NH–no hit.

know, the described study is the first application of comparing three datasets containing three-dimensional data, stored in different data types. Additionally, the created 3DTennisDS seems to collect the precise data that can further be applied in the classification purposes.

CONCLUSIONS

In this paper the new state-of-the-art 3DTennisDS dataset has been presented. This is the first tennis database publicly available that contains three-dimensional data of four types of strokes captured using the Vicon optical Mocap system applying the Pig model. The recorded movements are as follow: backhand, forehand and volleys. Apart from the athlete's silhouette, the created dataset also involves the model of tennis racket consisting of seven markers, which is an additional benefit in classification process. Each tennis stroke is stored in a separate c3d file, so they are ready for further use. In order to verify the potential of the created 3DTennisDS dataset, a series of experiments have been performed with the ST-GCN classifier. Our dataset has been compared to the very well-known THETIS dataset and to the Tennis-Mocap dataset in the field of tennis movements classification. These three datasets contain tennis strokes represented in various ways. The THETIS dataset was collected using a markerless motion capture system, while the Tennis-Mocap dataset was recorded using a marker-based motion capture system represented in the form of bvh structure. The last one, presented in this paper, was recorded with the optical Mocap system using the PiG model, which data are stored in the form of c3d files.

The classification of four main tennis strokes was performed. Based on the obtained results, higher accuracy for tennis action recognition has been achieved for the newly created dataset, 3DTennisDS. It can be stated that the way of capturing data and its precision has a great impact on the classification results. What is more, it turned out that data fuzzyfication has a positive impact on classification performance for analysed datasets. Applying this method improves the accuracy for tennis strokes recognition.

Due to the need for accurate data that represent various sport activities as well as the obtained classification results, it can be concluded that the created 3DTennisDS may have an impact on the development of scientific research and the publication of new articles.

Acknowledgements

The research program titled "Biomechanical parameters of athletes in individual exercises" and based on the analysis of 3D motion data and EMG, realized in the Laboratory of Motion Analysis and Interface Ergonomics was approved by the Commission for Research Ethics, No. 2/2016 dated 8.04.2016. The authors would like to thank the Tennis Academy POL-STAR Student Sports Club KS-WINNER Club for their support. The 3DTennisDS dataset is publicly available at https://tennisdb.cs.pollub.pl/.

REFERENCES

1. Rajšp A., Fister Jr, I. A systematic literature review of intelligent data analysis methods for smart sport training. Appl. Sci. 2020, 10(9): 3013.
2. Tennis Creative. Tennis Popularity Statistics 2021. https://tenniscreative.com/tennis-popularity-statistics/, 2022, [Online: 9.05.2024].
3. Vincent J., Stergiou P., Katz L. The role of databases in sport science: Current practice and future potential. International Journal of Computer Science in Sport. 2009, 8(2): 50–66.
4. Zhang W., Liu Z., Zhou L., Leung H., Chan A. Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation. Image Vis. Comput. 2017, 61: 22–39.
5. Gourgari S., Goudelis G., Karpouzis K., Kollias S. Thetis: Three dimensional tennis shots a human action dataset. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013, 676–681.
6. Cai J., Hu J., Tang X., Hung T.-Y., Tan Y.-P. Deep historical long short-term memorys for action recognition. Neurocomputing. 2020, 407: 428–438.
7. Sen A., Hossain S. M., Uddin R., Deb K, Jo K.-H. Sequence recognition of indoor tennis actions using transfer learning and long short-term memory. In: International Workshop on Frontiers of Computer Vision, Springer, 2022, 312–324.
8. Mora S.V., Knottenbelt W.J. Deep learning for domain-specific action recognition in tennis. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, 170–178.
9. Mora S. Vision and machine learning for in-play tennis analysis: framework, algorithms and implementation. Imperial College of London, Technology and Medicine, 2017.
10. Ullah M., Mudassar Yamin M., Mohammed A., Daud Khan S., Ullah H., Alaya Cheikh F. Attention-based lstm network for action recognition in sports. J. Electron. Imaging. 2021, 33: 1–6.
11. Vainstein J. Manera J., Negri P., Delrieux C., Maguitman A. Modeling video activity with dynamic phrases and its application to action recognition in tennis videos. In: n Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2–5, 2014. Proceedings 19, 909–916. Springer International Publishing.
12. Tsatiris G., Karpouzis K., Kollias S. Variance-based shape descriptors for determining the level of expertise of tennis players. In: 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). 2017, 169–172.
13. Chang H. Research on tennis motion evaluation method and system based on deep learning. Int T Electr Energy. 2022.
14. Valencia-Marin C. K., Pulgarin-Giraldo J. D., Velasquez-Martinez L. F., Alvarez-Meza A. M,. Castellanos-Dominguez G. An enhanced joint Hilbert embedding-based metric to support mocap data classification with preserved interpretability. Sensors. 2021, 21(13): 4443.
15. Yamada K., Matsuura M., Hamagami K., Inui H. Motor skill development using motion recognition based on an hmm. Procedia Comput. Sci. 2013, 22: 1112–1120.
16. Bačić B. Towards the next generation of exergames: Flexible and personalised assessment-based identification of tennis swings. In: 2018 International Joint Conference on Neural Networks. IEEE. 2018, 1–8.
17. Bezobrazov S., Sheleh A., Kislyuk S., Golovko V., Sachenko A., Komar M., Dorosh V., Turchenko V. Artificial intelligence for sport activitity recognition. In: 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IEEE. 2019, 2: 628–632.
18. Ganser A., Hollaus B., Stabinger S. Classification of

tennis shots with a neural network approach. Sensors. 2021, 21: 17.

19. Ma K. A real time artificial intelligent system for tennis swing classification. In: 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE. 2021, 000 021–000 026.
20. Ó Conaire C., Connaghan D., Kelly P., O'Connor N. E., Gaffney M., Buckley J. Combining inertial and visual sensing for human action recognition in tennis. In: Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams. 2010, 51–56.
21. Skublewska-Paszkowska M., Powroznik P., Lukasik E. Learning three dimensional tennis shots using graph convolutional networks. Sensors. 2020, 20(2): 6094.
22. Weinzaepfel P., Rogez G. Mimetics: Towards understanding human actions out of context. Int. J. Comput. Vis. 2021, 129: 1675–1690.
23. Kviatkovsky I., Rivlin E., Shimshoni I. Online action recognition using covariance of shape and motion. Comput. Vis. Image Underst. 2014, 129: 15–26.
24. Soomro K., Zamir A., Shah M. Ucf101: A dataset of 101 human actions classes from videos in the wild. CoRR. 2012, 12.
25. López-Méndez A., Alcoverro M., Pardàs M., Casas J.R. Approximate partitioning of observations in hierarchical particle filter body tracking. In: Cvpr 2011 Workshops. IEEE. 2011, 19–24.
26. Matsugu M., Yamanaka M., Sugiyama M. Detection of activities and events without explicit categorization. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). 2011, 1532–1539.
27. Tsagkatakis G., Jaber M., Tsakalides P. Convolutional neural networks for the analysis of broadcasted tennis games. Electron. Imagin. 2018, 206–2011.
28. Kurose R., Hayashi M., Aoki Y. Estimation of scenes contributing to score in tennis video using attention. In: 2018 14th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS). 2018, 672–675.
29. Sukhwani M. Jawahar C. Frame level annotations for tennis videos. In: 2016 23rd International Conference on Pattern Recognition (ICPR). 2016, 841–846.
30. Sukhwani M. Jawahar C.V. Tennisvid2text: Fine-grained descriptions for domain specific videos. ArXiv. 2015, abs/1511.08522.
31. Shimizu T., Hachiuma R., Saito H., Yoshikawa T., Lee C. Prediction of future shot direction using pose and position of tennis player. 2019, 10: 59–66.
32. Connaghan D., O'Conaire C., Kelly P., O'Connor N. E. Recognition of tennis strokes using key postures. In: IET Irish Signals and Systems Conference (ISSC 2010). 2010, 245–248.
33. Kobayashi S., Kaseda H., Miyamoto R. Robust localization of body parts based on interframe failure correction. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics. 2018, 3903–3908.
34. Faulkner H., Dick A. Tenniset: A dataset for dense fine-grained event recognition, localisation and description. In: 2017 International Conference on Digital Image Computing: Techniques and Applications. 2017, 1–8.
35. Huang Y.-C., Liao I.-N., Chen C.-H., Ik T.-U., Peng W.-C. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications. In: 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE. 2019, 1–8.
36. Maheswari U., Ramakrishnan R. Sports video classification based on the distribution of edge strengths of non subsampled shearlet transform. Int. J. Appl. Eng. Res. 2014, 9(23): 21835–21850.
37. Ely J., Gauriot R., Page L. Do agents maximise? risk taking on first and second serves in tennis. J. Econ. Psychol. 2017, 63: 135–142.
38. Wei X., Lucey P., Morgan S., Carr P., Reid M., Sridharan S. Predicting serves in tennis using style priors. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015, 2207–2215.
39. Qazi T., Mukherjee P., Srivastava S., Lall B., Chauhan N. R. Automated ball tracking in tennis videos. In: 2015 Third International Conference on Image Information Processing. 2015, 236–240.
40. Wu J., Liu D., Guo Z., Xu Q., Wu Y. Tacticflow: Visual analytics of ever-changing tactics in racket sports. IEEE Trans. Vis. Comput. Graph. 2022, 28(1): 835–845.
41. FarajiDavar N., De Campos T., Kittler J., Yan F. Transductive transfer learning for action recognition in tennis games. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE. 2011, 1548–1553,
42. Zhu G., Xu C., Huang Q., Gao W., Xing L. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. In: Proceedings of the 14th ACM international conference on Multimedia. 2006, 431–440.
43. Zhu G., Xu C., Gao W., Huang Q. Action recognition in broadcast tennis video using optical flow and support vector machine. In: Computer Vision in Human-Computer Interaction: ECCV 2006 Workshop on HCI, Graz, Austria, Proceedings 9. Springer, 2006, 89–98.
44. Zhu G., Xu C., Huang Q., Gao W. Action recognition in broadcast tennis video. In: 18th International Conference on Pattern Recognition. 2006, 1: 251–254. IEEE.
45. Davis R. B., Ounpuu S., Tyburski D., Gage J. R. A gait analysis data collection and reduction technique.

Hum. Mov. Sci. 1991, 10(5): 575–587.

46. Yan S., Xiong Y., Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Thirty-second AAAI conference on artificial intelligence. 2018.
47. Dai J., Qi H., Xiong Y., Li Y., Zhang G., Hu H., Wei Y. Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision. 2017, 764–773.
48. Kipf T. N., Welling M. Semi-supervised classification with graph convolutional networks. arXiv. 2016, preprint arXiv, 1609.02907.
49. Duddu V., Samanta D., Rao D.V. Fuzzy graph modelling of anonymous networks. In: International Workshop Soft Computing Applications. Springer. 2018, 432–444.
50. Krleža D., Fertalj K. Graph matching using hierarchical fuzzy graph neural networks. IEEE Trans. Fuzzy Syst. 2016, 25(4): 892–904.
51. Minaee S., Boykov Y.Y., Porikli F., Plaza A.J., Kehtarnavaz N., Terzopoulos D. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2021.
52. Pei W., Wang J., Xu X., Wu Z., Du X. An embedded 6-axis sensor based recognition for tennis stroke. In: 2017 IEEE International Conference on Consumer Electronics. IEEE. 2017, 55–58.
53. Skublewska-Paszkowska M., Powroznik P., Lukasik E. Attention temporal graph convolutional network for tennis groundstrokes phases classification. In: 2022 IEEE International Conference on Fuzzy Systems. IEEE. 2022, 1–8.