Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal 2023, 17(5), 313–325
https://doi.org/10.12913/22998624/171592
ISSN 2299-8624, License CC-BY 4.0

Received: 2023.07.03
Accepted: 2023.08.27
Published: 2023.10.20

Dual Attention Graph Convolutional Neural Network to Support Mocap Data Animation

Maria Skublewska-Paszkowska[1], Pawel Powroźnik[1*],
Marcin Barszcz[1], Krzysztof Dziedzic[1]

[1] Faculty of Electrical Engineering and Computer Science, Department of Computer Science, Lublin University of Technology, ul. Nadbystrzycka 38 D, 20-618 Lublin, Poland

* Corresponding author's e-mail: p.powroznik@pollub.pl

ABSTRACT

The analysis of movements is one of the notable applications within the field of computer animation. Sophisticated motion capture techniques allow to acquire motion and store it in a digital form for further analysis. The combination of these two aspects of computer vision enables the presentation of data in an accessible way for the user. The primary objective of this study is to introduce an artificial intelligence-based system for animating tennis motion capture data. The Dual Attention Graph Convolutional Network was applied. Its unique approach consists of two attention modules, one for body analysis and the other for tennis racket alignment. The input to the classifier is a sequence of three dimensional data generated from the Mocap system and containing an object of a player holding a tennis racket and presenting fundamental tennis hits, which are classified with great success, reaching a maximum accuracy over 95%. The recognised movements are further processed using dedicated software. Movement sequences are assigned to the tennis player's 3D digital model. In this way, realistic character animations are obtained, reflecting the recognised moves that can be further applied in movies, video games and other visual projects.

Keywords: Tennis strokes recognition, graph networks, convolutional networks, computer animation.

INTRODUCTION

Nowadays, there is a growing demand for precise data reflecting three-dimensional movement, especially for the needs of biomechanics [1], medical applications, computer games, computer animation [2, 3], sport, and activity recognition [4]. Data obtained from various motion capture systems (Mocap) contain the human and object pose expressed as a multi-channel time-based data. In each frame sequences of corresponding 3D points are stored [5]. Sophisticated equipment, like depth sensors or optical cameras operating in near infrared, allows to acquire the accurate data. Such precise 3D data represent variations in actions, and can further be applied in recognition, classification or animation. The skeleton-based action recognition is a challenging task. Its first step is to learn the sequences of the given points in 3D data and finally identify unique patterns [5]. From this type of feature detection, extracted e.g., the extraction process involves obtaining local spatio-temporal occupancy patterns or a 3D scene flow from 3D skeletal joints, silhouettes, and body part locations [6, 7]. Based on them the classification process is performed.

Recognising individual or team actions in sport has an essential function in determining the players' performance level as well as evaluating training. Various deep neural classification methods have been applied in basketball [8], volleyball [9], karate athletes [10] and watersports [11]. Based on the results obtained in [12], the temporal neural networks were deemed to be more appropriate for human action recognition. Graph convolutional neural networks have often been used in sport action recognition because of existing patterns and dependencies embedded in spatial configuration of human joints [13-15]. In some sports, additional objects are helpful in recognising movements. An example of such a sport is tennis, where the arrangement of the tennis racket makes it easier to identify moves [16].

Tennis movement classification using motion capture data has been considered in many studies. Based on data recorded by Kinect, twelve tennis strokes (various types of forehand, backhand, volley and serve) were recognized using the following approaches: Long Short Term Memory (LSTM) artificial model with 3 layers [17, 18] and five ones [20], SVM, and Conditional Random Fields (CRF) with a linear chain [19]. Another study concerning tennis strokes, described in [21], is about temporal location extraction utilized cost-effective visual and inertial sensing. The classification of tennis moves was performed with the use of SVM and kNNs. Classification based on boost Hilbert using embedding approach with a cross-correlation operator based on Tennis-Mocap and HDM05 was proposed in [2]. Serve, forehand, backhand, volley, backhand volley and smash were recognized.

Spatial Temporal Graph Convolutional Networks were applied with great success for forehand, backhand, and no-shot in [22, 23]. Another approach, reaching satisfactory results, involving the Attention Temporal Graph Convolutional Network (A3T-GCN) was proposed in [24]. Movements of various sports were animated for the purposes of teaching or verification. Animation performed on the basis of motion capture data to control virtual world was presented in [25]. A deep model, called Cool-TSN, was applied for action recognition. A 3D table tennis simulation animation for the purpose of teaching was described in [26, 27]. The animation of tennis moves of a professional player versus a trainee was presented in [28]. Eleven forehand strokes were compared. Computer graphics animation of a tennis serve involving a whole body and a racket based on video capturing for the purpose of judgement of ball direction was presented in [29]. In the study 15 professional and 15 amateur participants took part.

Point-light motion animations of forehand groundstrokes, both for left-handed and right-handed, based on a marker-based motion capture system was presented in [30]. Twenty professional and twenty non-players participated in the study. A VR system, in which the simulated player was rendered in a real time utilizing implementing artificial intelligence, dedicated for sport education, was described in [31]. Various approaches were applied, such as intelligent animation control, stereoscopic display with high-definition, hybrid tracking, haptics feedback, skin deformation based on shades as well as VR immersive experience in a real time. The backward, forward, and sideways motions toward the ball were applied.

The main contribution of this paper is to propose a system for tennis motion capture data animation involving artificial intelligence. For human movement recognition, the Dual Attention Graph Convolutional Network (DA-GCN), was presented. Its unique approach consists of two attention modules, one for body analysis and the other for tennis racket alignment. The network takes into account both spatial and temporal features. It involves GCN and LSTM. The classifier is trained with a set of 3D data as input. The data are generated from the Mocap data and containing an object of a player holding a tennis racket. Based on this type of information, tennis backhand, forehand, and two types of volleys have been classified with great success. The recognised movements are further processed using Autodesk's MotionBuilder software. Movement sequences are assigned to the tennis player's 3D digital model. In this way, realistic character animations were obtained that reflected the recognised moves. This allows to export them to various file formats compatible with various rendering programs. Therefore, they can be used in movies, video games and other visual projects. To the best to the author's understanding this study presents the first attempt to create animation using c3d files for tennis data presenting a tennis player together with a racket.

MOCAP DATA

Data acquisition

The recordings were completed in an indoor laboratory. The experiment involved three female and seven male tennis players performing backhand, forehand, volley backhand and volley forehand strokes. The movements were recorded using an advanced eight-camera T40S optical Vicon motion capture system, operating at a frequency of 100 Hz.

The recordings were completed in an indoor laboratory. To add a natural element to the strokes, the backhand and forehand moves were executed while the participants were running and navigating around an obstacle placed on the floor. This setup facilitated more realistic strokes compared to hitting the ball from a stationary position. Initially, each participant performed ten forehand strokes without a ball, followed by ten backhand strokes without a ball. These exercises were then repeated, incorporating a ball. At the end, the participants executed ten volleys forehand and ten volleys backhand in front of a tennis net, with tennis balls being thrown from both the right and left sides of the net. While standing parallel to the net, the players made short movements with their rackets in front of them, causing the balls to bounce and fall.

Participant preparation

The study involved a total of ten tennis players, each of them played as a professional player at least 7 years. They were qualified based on their experience. They underwent preparation according to the Plug-in Gait specifications. To accurately capture their movements, thirty-nine retroreflective markers were attached to the players' bodies. Additionally, an extra tennis racket model was prepared specifically for registration purposes, and it was equipped with seven retroreflective markers. These markers were positioned on the top and bottom of the racket head, and at the bottom of the racket handle. Also two markers were attached on both sides of the racket.

Data post-processing

All incorrectly-performed strokes were rejected. All other recordings were post-processed according to the following procedure. Firstly, each marker was labelled conforming to the Plug-in Gait model or tennis racket model description. Secondly, all missing markers were interpolated in order to ensure the continuity of the trajectory. Thirdly, all recordings were cleaned so that there were only markers with label assigned. Finally, the Plug-in Gait model was applied for thirty-nine markers, reflecting a tennis player. The obtained data were exported using c3d format.

It should be noted that all registered data were carefully checked by a tennis specialist. Totally, the final dataset included: 212 backhand, 197 forehand, 180 forehand volley and 180 backhand volley.

DA-GCN CLASSIFIER

In order to verify the study assumptions, a Dual Attention Graph Convolutional Neural Network (DA-GCN) was introduced. Its architecture was presented in Figure 1. Proposed DA-GCN consisted of four main elements connected with: spatial and temporal features extraction, attention module as well as fully connected layers. Classifier input was presented in the form of a graph $H=(Z,D)$, where Z indicates the graph vertices while D denotes its edges. The proposed input consisted of N nodes, connected with N skeleton joints and their altering in time t. The final structure of node could be defined as $Z=\{z_{ti}|t=0,...,T,\ i=1,...,N\text{-}1\}$. Spatial and temporal features were obtained from the three-dimensional data in order to obtain the most accurate prediction of tennis movement. Two attention modules have been implemented to improve the accuracy of tennis stroke classification. The first one was responsible for the unambiguous location of the player's silhouette, while the second for the position of the tennis racket. The applied classifier will enable the most accurate representation of the tennis player's realistic movement.

In the proposed solution, the graph convolution operation (Eq. 1) is applied directly to the input data. This allows the extraction of highly significant patterns and features in the space domain. The graph convolution operator "*G" is defined as the multiplication of a signal $X \in R^n$ with a kernel Θ.

$$\Theta * \mathrm{G}\,x = \Theta(L)X = \Theta(U\Lambda U^T)X = U\Theta(\Lambda)U^T X \quad (1)$$

where: U denotes the matrix of eigenvectors of the normalized graph Laplacian L, $U^T X$ is a Fourier Transform, Λ indicates the diagonal matrix of eigenvalues of L.

Following the completion of graph convolution operations that have harnessed neighboring information for each node within the spatial dimension of the graph, an additional standard convolution layer is sequentially applied along the temporal dimension. This layer serves to enhance the node's signal by amalgamating the insights from adjacent time periods. The attention module comprises a basic 2D-convolutional layer, followed by a sigmoid function that produces a mask for the input feature map. Operating on a three-dimensional feature map input, it yields a X×Y×T attention map output. This resultant attention map

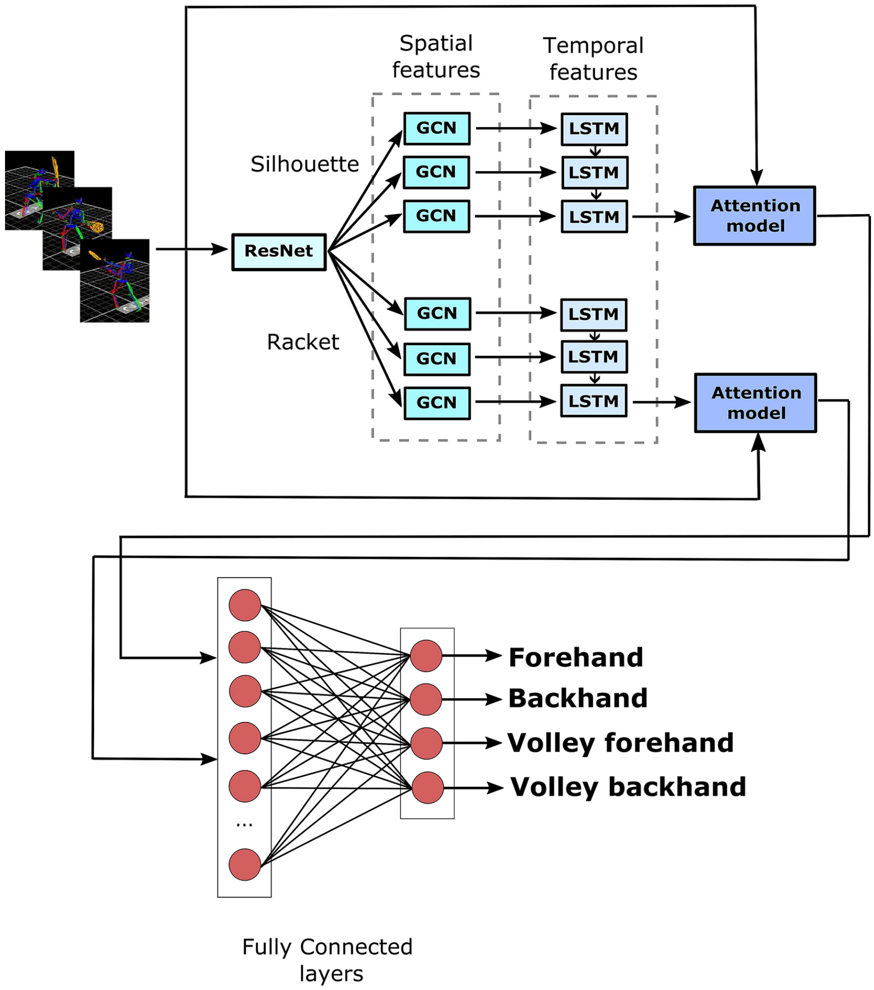

Fig. 1. General scheme of DA-GCN classifier

is subsequently multiplied with the input feature map through element-wise multiplication.

Spatial features

To extract spatial features from the acquired tennis strokes, graph convolutional networks (GCNs) were applied. GCNs allow to determine the spatial relationship between individual graph nodes, described by Eq. (2), due to used Fourier filter [33]:

$$B^{k+1} = \sigma\left(\widetilde{S_A}^{-\frac{1}{2}}\tilde{A}\widetilde{S_A}^{-\frac{1}{2}}B^{(k)}\psi^{(k)}\right) \tag{2}$$

where: $\tilde{A} = I_n + A$ – denotes adjacency matrix A with self-connections in a form of identity matrix I_n; $\widetilde{S_A}$ – a sum of adjacency matrix coefficients $(\widetilde{S_A} = \sum_i \widetilde{A_{Jl}})$. $B^{(k)}$ denotes the output of k^{th} layer, while $\psi^{(k)}$ store the parameters of k^t layer. The sigmoidal function is represented as σ for nonlinear model [33].

To extract spatial features for series of input skeleton graphs a GCN consisting of 3 layers was applied. Mathematically, for feature input X, it could be expressed in a way provided by Eq. (3):

$$f(X, A) = \sigma\left(\hat{A}ReLU\left(\left(\hat{A}XW_0\right)\hat{A}XW_1\right)W_2\right) \tag{3}$$

where: $\hat{A} = \widetilde{S_A}^{-\frac{1}{2}}\tilde{A}\widetilde{S_A}^{-\frac{1}{2}}$ indicates the pre-processing stage, W_0, W_1, W_2 – represents the weights for layers: input, hidden and output. Although the dimension of the weights layer is different, precisely $W_0 \in R^{Lxl}$, L – a feature matrix length and l – denotes number of inputs to the hidden layer W_1, $W_2 \in R^{0xl}$, O – an output size. *ReLU* – stands for the Rectified Linear Unit activation function.

Temporal features

In order to designate temporal features, which were one of the most important issues to classify tennis strokes, RNN networks were applied. Unfortunately, their typical form turned out to be insufficient, therefore LSTM was used, which in this case processes information based on three items of data related to the strokes at times t, t_{-1}, and t_{-2}. It should be emphasized that in order to determine time dependences for tennis player model, the knowledge of both the current moment of data and the previous one are needed. The applied architecture of the LSTM network was conditioned on the given Equations 4-6 [34]:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \tag{5}$$

$$f_o = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{6}$$

where: i_t, f_o, f_t represented the gates: input, output and forget, w_i, w_f, w_o were weights indicating individual gates, b_i, b_f, b_o denotes bias values for them, h_{t-1} was the value of the result of the earlier LSTM block at moment t-1, while x_t dentoted an input at present moment, and indicated the sigmiodal function. Furthermore, the formulas for the gate state could be described by Eq. 7– 9 [34]:

$$\widetilde{c_t} = tanh(w_c[h_{t-1}, x_t] + b_c) \tag{7}$$

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c_t} \tag{8}$$

$$t_h = o_t * \tanh(c_t) \tag{9}$$

where: c_t denoted the gate state at timestamp t, while $\tilde{c_t}$ represented the pretender for the gate state at the moment t.

Attention module

Extraction of discriminant features is necessary to capture contextual information. For this purpose, two attention modules have been proposed. They were used to capture information involving the tennis moves of both the player and the racket in the form of feature attention matrices. These elements were necessary to predict their future position. In Figure 2 the general structure of attention module was presented. As the stage, three-dimensional input features, size X×Y×T ∈ R, were transformed into the feature matrix size X×X ∈ R. To obtain it the operation of matrix reshaping (RS) and transposition (TP) were performed. The final form was achieved based on the Softmax function. Moreover, the multiplication between input parameters and feature matrix was done. Furthermore, results were multiplied by scale parameter and the elementwise sum operation was calculated. The whole process of indicating the n feature map could be represented by Eq. 10:

$$O_n = \beta \sum_{k=1}^{X} (i_{nk} I_k) + I_n \tag{10}$$

where: O – represents the output value, i_{nk} – an impact of the k^{th} feature on n^{th}, I_n – denotes the input parameters and β is a weight parameter.

CLASSIFIER EVALUATION

In order to check the advisability of using classifier consisting of two attention models, preliminary tests were carried out. In this study, the fragments responsible for recognizing the silhouette and the racket (Fig. 1), was removed from proposed model, respectively. Two studies were conducted: one for the silhouette itself and the other for the racket itself. In Tables 1-2 the results of the obtained Accuracy measures are gathered.

Analyzing the results, it is obvious that the achieved results of individual components are far from acceptable. Due to it, the double attention classifier has been verified by the following measures: Accuracy, Precision, Recall and F1 score (see Tables 3-6). The confusion matrix for the analysed tennis strokes was calculated for indicating the misleading classification (Fig. 3). The learning curve and the loss function are shown in Figure 4.

The summary of the works concerning tennis strokes recognition is presented in Table 7. These state-of-the-art studies involve many various acquisition technique, including sensors, video and motion capture systems. A great number of the research in this domain has been conducted on the widely recognized THETIS database. Both video data and images derived from the Kinect motion capture system have been employed to analyze and identify tennis movements. Additionally, real match broadcasts have frequently been incorporated into the research. Various neural network approaches

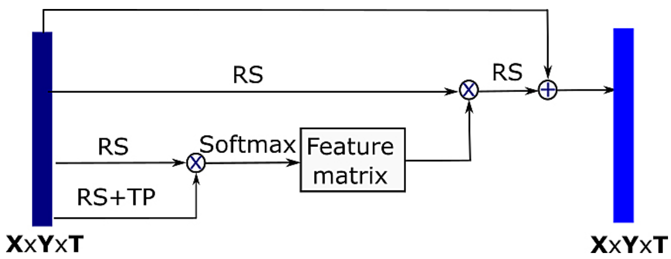

Fig. 2. General scheme of attention module. RS represents matrix reshaping, TP matrix transposition

Table 1. Obtained accuracy for forehand (F), backhand (B), volley forehand (VF) and volley backhand (VB) (only silhouette)

Stroke	Mean	Max	Min	±SD
F	70.94%	74.86%	65.51%	4.54%
B	70.13%	74.29%	65.38%	5.57%
VF	70.73%	76.20%	63.23%	5.35%
VB	71.41%	78.96%	63.21%	5.44%

Table 2. Obtained Accuracy for Forehand (F), Backhand (B), Volley Forehand (VF) and Volley Backhand (VB) (only racket)

Stroke	Mean	Max	Min	±SD
F	67.25%	72.89%	60.94%	7.03%
B	66.94%	72.96%	60.11%	6.35%
VF	67.03%	72.83%	60.53%	4.95%
VB	66.19%	72.20%	60.45%	7.39%

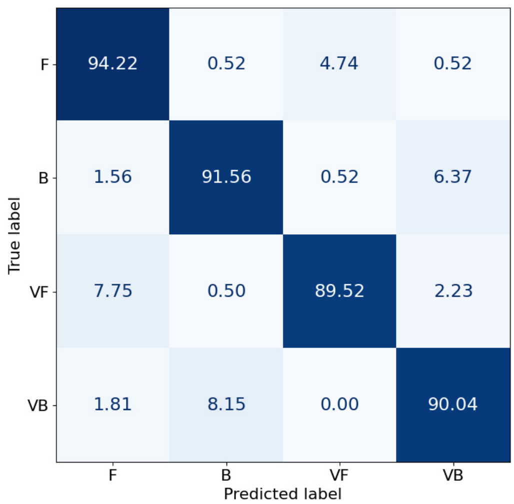

Fig. 3. Confusion matrix (in %) for forehand (F), backhand (B), volley forehand (VF) and volley backhand (VB)

have been utilized to achieve these objectives. Notably, graph neural networks, have been deployed to identify tennis movements using motion capture data. This type of network stems from the distinctive characteristics of the recorded data. In this approach, a human model is created with markers affixed to specific locations on the body. Based on this data a graph is created, which accurately represents the character of the human body.

TENNIS STROKES ANIMATION

Animating a digital 3D model using Autodesk's MotionBuilder software is one of the popular ways to create realistic animations of human figures and objects. Due to this software, it is possible to precisely reproduce the movements and behavior of the characters, which is especially useful in various areas including sports. In the described process, the data reflecting the movement of a real tennis player was saved in a file in .c3d format. Moreover, it was possible to import the data into the MotionBuilder software, in which the data could be used to animate the characters. The sequence of movement in the 3D scene was presented as a cloud of distinguishable animated points (Figure 5) which correspond

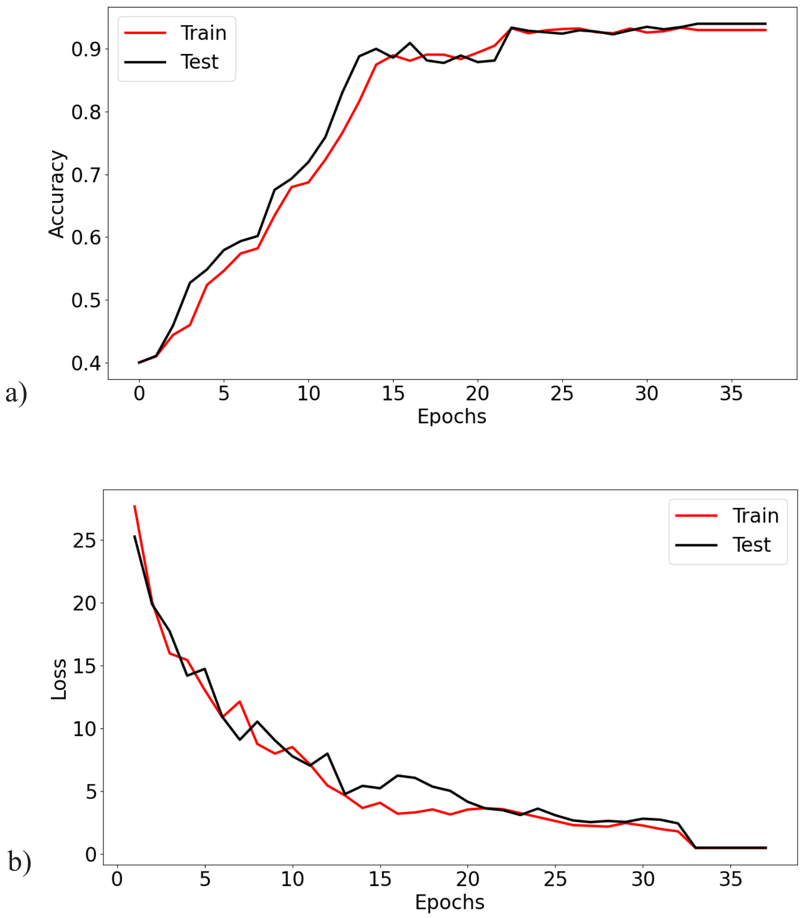

Fig. 4. Selected learning parameter

Table 3. Obtained accuracy for forehand (F), backhand (B), volley forehand (VF) and volley backhand (VB)

Stroke	Mean	Max	Min	±SD
F	91.38%	96.17%	86.00%	3.19%
B	91.13%	95.98%	85.01%	3.40%
VF	89.79%	95.48%	85.63%	3.31%
VB	90.41%	96.91%	86.74%	3.61%

Table 4. Obtained Precision for Forehand (F), Backhand (B), Volley Forehand (VF) and Volley Backhand (VB)

Stroke	Mean	Max	Min	±SD
F	97.63%	98.97%	89.58%	2.73%
B	94.34%	98.97%	87.76%	3.40%
VF	92.40%	97.96%	85.15%	3.80%
VB	92.01%	95.05%	86.74%	2.47%

Table 5. Obtained recall for forehand (F), backhand (B), volley forehand (VF) and volley backhand (VB)

Stroke	Mean	Max	Min	±SD
F	91.82%	96.97%	85.15%	3.45%
B	93.04%	96.94%	86.87%	3.06%
VF	97.60%	98.97%	90.53%	2.47%
VB	93.91%	98.97%	86.73%	3.26%

to the positions of the markers on the real tennis player. In order to create a realistic animation, anatomically matched virtual actor must be added to the point cloud, which will be set in motion (Figure 6a). For this purpose, it is necessary to create the so-called performer rig, i.e., an auxiliary skeleton that will be moved by a cloud of points. The rig consists of properly arranged bones that will control the character's movements. Based on it, the skeleton of the animated character will be created and coupled with the performer skeleton (Figure 6b). The last stage is the verification of the correctness of the animation and its fixation on the skeleton of the character (Figure 7). Due to this, the model will be able to move in the desired way without the participation of performer skeleton and markers.

Table 6. Obtained F1 for forehand (F), backhand (B), volley forehand (VF) and volley backhand (VB)

Stroke	Mean	Max	Min	±SD
F	94.62%	97.96%	87.31%	2.93%
B	93.69%	97.36%	87.31%	3.20%
VF	94.92%	98.49%	87.76%	3.02%
VB	92.95%	96.97%	86.74%	2.83%

Table 7. The comparison of tennis movement classification (NH – no hit, VB – volley backhand, VF – volley forehand, B – backhand, F – forehand, H – hit, BS – backspin, SM – smash, S – serve, V – volley)

Name of dataset/ type of data	Source input	Tennis moves	Neural approaches	Accuracy	Study
SensorTile	Signal	F, B, S	DNN	94.00-97.00%	[35]
IMU	Signal	F, B, S, BS, SM	SVM	90.85-98.86%	[36]
			NN	98.76-100.00%	
			DT	84.69-95.54%	
			RF	93.75-98.96%	
			kNN	87.76-99.44%	
THETIS	Video	B, V, F, S, SM	LSTM	81.23-89.42%	[37]
			SVM	51.20%	[19]
			CRF	86.44%	
THETIS	Video	B, V, F, S, SM	Deep Historical LSTM	62.00%	[20]
HMDB51				54.00%	
THETIS	Video	B, V, F, S, SM	LSTM	70.17-97.67%	[38]
KTH					
THETIS	Video	B, V, F, S, SM	SVM	53.08-60.23%	[39]
KTH				90.65	
KTH	Video	S, H, NH	KLDA	73.34-92.29%	[40]
Broadcast	Video	F, B	SVM	90.21%	[41]
				87.10%	[42, 43]
mix	signal, video	F, B, S	SVM	82.43-97.02%	[21]
			kNN	84.73-100.00%	
Vicon with fuzzy input	Images	F, B, NH	ST-GCN	86.30-87.30%	[22]
			A3T-GCN	86.90-93.82%	[24]
Vicon without fuzzy input			ST-GCN	64.10-74.30%	[22]
			A3T-GCN	74.22-81.95%	[24]
Vicon	3D silhouette	F, B, V	A3T-GCN	78.33-85.54%	[16]
	3D silhouette & racket			85.62-93.98%	[16]
Vicon	.c3d	F, B, VF, VB	DA-GCN	85.01-96.91%	This work

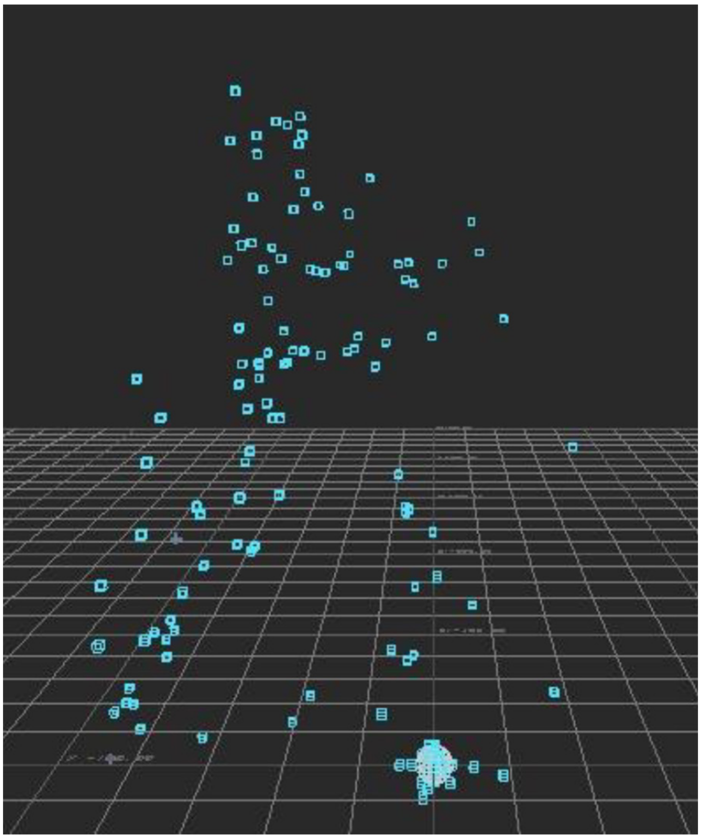

Fig. 5. Animated point cloud based on .c3d data

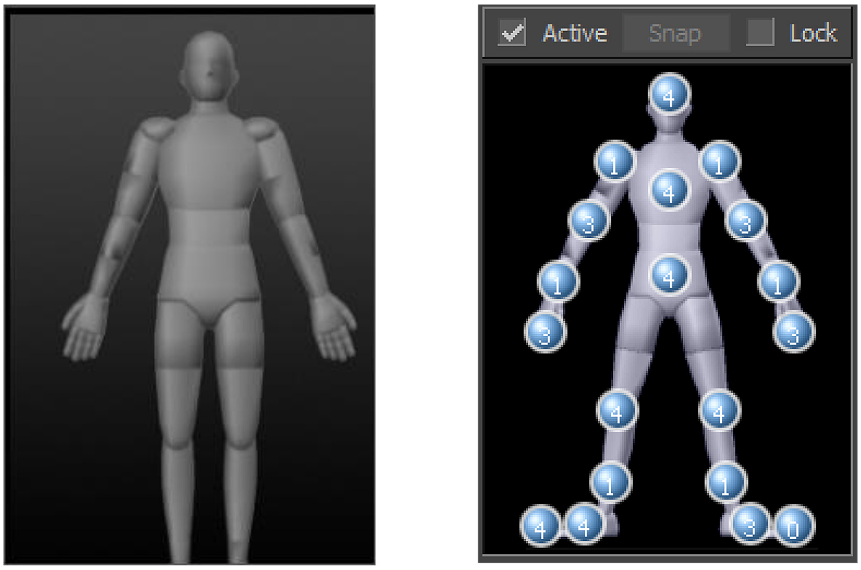

Fig. 6. (a) 3D model of an actor, (b) 3D model of an actor with markers

DISSCUSION

The DA-GCN classifier for tennis moves, like backhand, forehand, volleys (backand and forehand) recognition was proposed with great success. In order to verify its effectiveness, measures like Accuracy, Precision, Recall and F1 score were computed (Tables 3-6). The Accuracy obtained for all tennis moves was higher than 89.7%, which confirms the high efficiency of the suggested recognition model. The highest accuracy was reached for volley backhand – 91.41%. For forehand and backhand the accuracy is also very high, at the

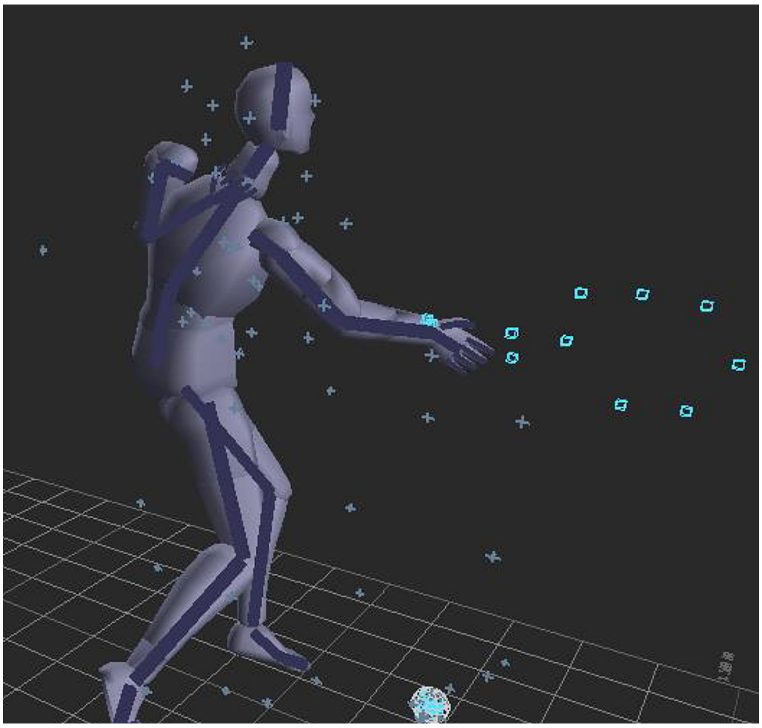

Fig. 7. Tennis player model based on .c3d data

level of 91%. It should also be indicated that all obtained maximum values exceeded 95.4%. Due to the unbalanced number of elements in the classes, Precision and Recall measures were calculated. They indicate classifier output quality. Precision determines the result relevance, while Recall informs about the number of truly returned relevant results. High results obtained for these two measures (average Precision higher than 92% and average Recall higher than 91.8%) indicate that classifier gains both accurate as well as positive results. F1 score is a measure for the model's accuracy evaluation. It specifies how often a correct prediction is performed taking into consideration the whole dataset. In this study the obtained average F1 score (Table 6) exceeds 92.9%, while the maximum value is higher than 96.9%. Analysing the confusion matrix, it can be observed that there is a tiny percentage of incorrectly classified data. The highest quantity of misrecognised strokes were for backhand instead of volley backhand (only 8.15%) and for forehand instead of volley forehand (only 7.75%). The confusion matrix also confirms that the proposed classification method is accurate for Mocap data.

Analyzing the state-of-the-art tennis movement recognition (Table 7), one can notice a limited number of studies based on optical motion capture systems [16, 22, 24]. It should be emphasized that the proposed solution, with its maximum effectiveness, exceeds the results obtained in other works [16, 22, 24]. One of the reasons for the classifier's behavior may be the use of a different type of input data extracted from the Vicon system. It should also be emphasized that the input data used in other works, such as: signals [35, 36], video obtained from the following databases: THETIS [19, 37, 38], KTH [38-40], HMD51 [20] whether the images [22, 24] are not as accurate as those used in this work. Furthermore deep neural networks [20] as well as less sophisticated [39-41] were also compared with proposed solution. Moreover, the proposed solution contains two models of attention that allow the extraction of features for the movement of the player and the tennis racket, separately. The structures used within the convolutional neural network (GCN and LSTM) enable independent separation of spatial and temporal features, which also affects the final classification efficiency. Nowadays, much attention is paid to the accuracy of mapping athletes' movements for the purpose of their analysis or the production of systems supporting novice athletes or coaches [26]. In order to reflect movements, spatial coordinates as well as the positions of the subject are needed. For this reason, more and more precise data are obtained from motion capture systems [27-28]. On the basis of the received trajectories, a data simulation may be created. Various animation techniques have been presented, such as Adobe Flash professional CS programming software [26], OpenGL [31] or CAD programs [44]. It is worth mentioning that in the group of racket sports simulations for table tennis [26-27] as well as solutions for tennis [28-30] have been developed. In the field of the last sport, the serving animation was developed, as well as the detection of the position of the tennis ball [29]. Selected tennis moves with a marked trajectory are presented in [28]. They did not contain a tennis racket, only the silhouette of a player. In [30], the visit of tennis players in the position of opponent was examined. The system presents an animation of right- and left-handed players in the form of markers reproducing the shape of the silhouette and the racket. This paper presents a novel system that combines the use of artificial intelligence to classify moves such as forehand, backhand, forehand volley and backhand volley. The modern techniques were involved. The new classification method was based on accurate 3D data recorded using the Vicon motion capture system. MotionBuilder was utilized for animation based on three-dimensional data. Whole body together with tennis racket was animated. Based on the literature review, it can be noticed that such a combination of data analysis and animation tools is a kind of novelty in scientific work.

CONCLUSIONS

The main purpose of the study was to create a software for the animation of classified basic tennis strokes. The experiment consists of two stages. Firstly, the classifier was proposed for tennis movement recognition. Secondly, the animation for these moves was applied. For the purposes of this study, a new model was built, consisting of graph convolutional neural networks as well as LSTM networks along with two attention mechanisms, allowing to clearly classify tennis strokes with accuracy exceeding other tools operating on similar data. The process of animating a 3D digital model using Autodesk's MotionBuilder software is a complex and requires precision as well as virtual character animation skills. However, this process allows one to create realistic character animation, which is applicable in many sport fields. In this study an animation using c3d Mocap data was performed for recognised tennis strokes. It presents the movements of a tennis player together with associated racket. The created system is a first attempt to build a sophisticated tool for virtual movement analysis as well as a learning software. Directions for future research may include extending the Mocap data with other tennis strokes, as well as the use of other animation methods.

Acknowledgement

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.

REFERENCES

1. Kadu H., Kuo C.C.J. Automatic human mocap data classification. IEEE Transactions on Multimedia 2014; 16(8): 2191-2202.
2. 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.
3. Shi M., Aberman, K., Aristidou A., Komura T., Lischinski D., Cohen-Or D., Chen B. Motionet: 3d human motion reconstruction from monocular video with skeleton consistency. ACM Transactions on Graphics (TOG) 2020; 40(1): 1-15.
4. Derlatka M., Borowska M., Ensemble of heterogeneous base classifiers for human gait recognition. Sensors 2023; 23(1): 508.
5. Kumar M.T.K., Kishore P.V.V., Madhav B.T.P., Kumar D.A., Kala N.S., Rao K.P.K., Prasad B. Can Skeletal Joint Positional Ordering Influence Action Recognition on Spectrally Graded CNNs: A Perspective on Achieving Joint Order Independent Learning. IEEE Access 2021; 9: 139611-139626.
6. Aggarwal J.K., Xia L. Human activity recognition from 3d data: A review. Pattern Recognition Letters 2014; 48: 70-80.
7. Sun Z., Ke Q., Rahmani H., Bennamoun M., Wang G., Liu J. Human action recognition from various data modalities: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 2022; 45(3): 3200-3225.
8. Ma C., Fan J., Yao J., Zhang T. NPU RGB+ D Dataset and a Feature-Enhanced LSTM-DGCN Method for Action Recognition of Basketball

Players. Applied Sciences 2021; 11(10): 4426.

9. Qi M., Wang Y., Li A., Luo J. Sports video captioning via attentive motion representation and group relationship modeling. IEEE Transactions on Circuits and Systems for Video Technology 2019; 30(8): 2617-2633.
10. Guo J., Liu H., Li X., Xu D., Zhang Y. An attention enhanced spatial–temporal graph convolutional LSTM network for action recognition in Karate. Applied Sciences 2021; 11(18): 8641.
11. Martinez B., Modolo D., Xiong Y., Tighe J. Action recognition with spatial-temporal discriminative filter banks. In: Proc. of the IEEE/CVF international conference on computer vision, Seul, Korea 2019, 5482-5491.
12. Nan M., Trăscău M., Florea A.M., Iacob C.C. Comparison between recurrent networks and temporal convolutional networks approaches for skeleton-based action recognition. Sensors 2021; 21(6): 2051.
13. Yan S., Xiong Y., Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proc. of the AAAI conference on artificial intelligence, New Orleans, USA 2018, 32 (1).
14. Chen Y., Tian Y., He M. Monocular human pose estimation: A survey of deep learning-based methods. Computer Vision and Image Understanding 2020, 192, #102897.
15. Liu J., Che Y. Action recognition for sports video analysis using part-attention spatio-temporal graph convolutional network. Journal of Electronic Imaging 2021; 30(3): 033017-033017.
16. Skublewska-Paszkowska M., Powroznik P. Temporal Pattern Attention for Multivariate Time Series of Tennis Strokes Classification. Sensors 2023; 23(5): 2422.
17. Mora S.V., Knottenbelt W.J. Deep learning for domain-specific action recognition in tennis. In: Proc. of the IEEE conference on computer vision and pattern recognition workshops, Honolulu, HI, USA, 2017, 114-122.
18. Mora S. V. Computer Vision and Machine Learning for In-Play Tennis Analysis: Framework, Algorithms and Implementation, Doctoral dissertation, Imperial College London, 2017.
19. Vainstein J., Manera J.F., Negri P., Delrieux C., Maguitman A. Modeling video activity with dynamic phrases and its application to action recognition in tennis videos. In: Proc. of 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico 2014, 909-916.
20. Cai J., Hu J., Tang X., Hung T.Y., Tan Y.P. Deep historical long short-term memory network for action recognition. Neurocomputing 2020; 407: 428-438.
21. 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: Proc. of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams, Firenze, Italy 2010, 51-56.
22. Skublewska-Paszkowska M., Powroznik P., Lukasik E. Learning three dimensional tennis shots using graph convolutional networks. Sensors 2020; 20(21): 6094.
23. Skublewska-Paszkowska M., Powroznik P., Karczmarek P., Lukasik E. Aggregation of Tennis Groundstrokes on the Basis of the Choquet Integral and Its Generalizations. In: Proc. of the 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Padua, Italy 2022.
24. Skublewska-Paszkowska M., Powroznik P., Lukasik E. Attention Temporal Graph Convolutional Network for Tennis Groundstrokes Phases Classification, In: Proc. of the 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Padua, Italy 2022.
25. de Souza C.R., Gaidon A., Cabon Y., Murray N., López A.M. Generating human action videos by coupling 3D game engines and probabilistic graphical models. International Journal of Computer Vision 2020; 128(5): 1505-1536.
26. Wang H. Research on design and implementation of computer 3D table tennis simulation animation. Journal of Physics: Conference Series 2021; 1744(3): 032088.
27. Yang X. Table Tennis Animation Simulation Aided Teaching System based on Box Particle Filter Tracking. In: Proc. of the 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India 2021, 926-929.
28. Oshita M. Motion Volume: Visualization of Human Motion Manifolds. In: Proc. of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, Brisbane QLD, Australia 2019, 1-7.
29. Ida H., Fukuhara K., Ishii M., Inoue T. Anticipatory judgements associated with vision of an opponent's end-effector: An approach by motion perturbation and spatial occlusion. Quarterly Journal of Experimental Psychology 2019; 72(5): 1131-1140.
30. Loffing F., Sölter F., Hagemann N., Strauss B. On-court position and handedness in visual anticipation of stroke direction in tennis. Psychology of Sport and Exercise 2016; 27: 195-204.
31. Song P., Xu S., Fong W.T., Chin C.L., Chua G.G., Huang Z. An immersive VR system for sports education. IEICE TRANSACTIONS on Information and Systems 2012; 95(5): 1324-1331.
32. Chen F., Pan S., Jiang J., Huo H., Long G. DAGCN: dual attention graph convolutional networks. In: Proc. of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary 2019, 1-8.
33. Zhao L., Song Y., Zhang C., Liu Y., Wang P., Lin T., Li H. T-gcn: A temporal graph convolutional network for traffic prediction. IEEE transactions on intelligent transportation systems 2019; 21(9): 3848-3858.

34. Bai J., Zhu J., Song Y., Zhao L., Hou Z., Du R., Li H. A3t-gcn: Attention temporal graph convolutional network for traffic forecasting. ISPRS International Journal of Geo-Information 2021; 10(7).
35. Ganser A., Hollaus B., Stabinger S. Classification of tennis shots with a neural network approach. Sensors 2021; 21(17).
36. Ma K. A real time artificial intelligent system for tennis swing classification. In: Proc of 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia 2021.
37. Pei W., Wang J., Xu X., Wu Z., Du X. An embedded 6-axis sensor based recognition for tennis stroke. In: Proc. of 2017 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV 2017, 55-58.
38. Ullah M., Mudassar Yamin M., Mohammed A., Daud Khan S., Ullah H., Alaya Cheikh F. Attention-based LSTM network for action recognition in sports. Electronic Imaging 2021; 2021(6): 302-1.
39. Gourgari S., Goudelis G., Karpouzis K., Kollias S. Thetis: Three dimensional tennis shots a human action dataset. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, Oregon, USA 2013, pp. 676-681.
40. FarajiDavar N., De Campos T., Kittler J., Yan F. Transductive transfer learning for action recognition in tennis games. In: Proc. of 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain 2021, 1548-1553.
41. 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: Proc. of the 14th ACM international conference on Multimedia, anta Barbara, CA, United States 2006, 431-440.
42. Zhu G., Xu C., Gao W., Huang Q. Action recognition in broadcast tennis video using optical flow and support vector machine. In: Proc. of the Computer Vision in Human-Computer Interaction: ECCV 2006 Workshop on HCI, Graz, Austria, 2006, 89-98.
43. Zhu G., Xu C., Huang Q., Gao W. Action recognition in broadcast tennis video. In: Proc. of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China 2006, vol. 1, 251-254.
44. Dziedzic K., Barszcz M., Paśnikowska-Łukaszuk M., Jankowska A. The role of computer animation in teaching technical subjects. Advances in Science and Technology Research Journal 2015; 9(28): 134-138.