Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 175 (2020) 149–156

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 9-12, 2020, Leuven, Belgium

iKarate: Karate Kata Guidance System

Bassel Emad[a,*], Omar Atef[a], Yehya Shams[a], Ahmed El-Kerdany[a], Nada Shorim[a], Ayman Nabil[a], Ayman Atia[b,c]

[a]*Faculty Of Computer Science, Misr International University (MIU), Cairo, Egypt*
[b]*Faculty Of Computer Science, October University for Modern Sciences and Arts (MSA), Cairo, Egypt*
[c]*Faculty Of Computer Science And Artificial Intelligence, HCI-Lab, Helwan University (HU), Cairo, Egypt*

Abstract

Smart coaching in martial arts is one of the recent research areas in Human Motion Analysis. Numerous moves are performed incorrectly during the performance. In this paper we offer a system that will record the Players' movements using IR (Infrared) camera sensor, store the data in a database, pre-process the data, classify the data using F-DTW (Fast Dynamic Time Warping) and then show the users an accurate report that contains every movement the player had done, their mistakes and how to improve their performance the next time. This approach focuses on the first seven movements of Karate Kata 1 (Hein Shodan). The system has reached an accuracy of 91.07% in classifying the moves and one common mistake for each move.

Keywords: Body Gestures; Classification; Computer Vision; Human Activity Recognition; Human Motion Analysis; Kinect; Motion Detection; Signal Analysis

1. Introduction

Karate Kata is a martial art that can be practiced individually or in groups. Karate Kata consists of sequences of movements done in a certain order and a certain motioning. Its performance quality is weighted by how well the player replicates the original move, which is difficult for the players to master. Children may have difficulties learning Karate Kata, since the training session consists of a wide number of players that makes the coach unable to observe all the details of every movement done by all the players. The lack of concentration with all the players may lead to taking more time to master the movements. In addition, some players would like to practice at home, especially in the current events caused by COVID-19 (Coronavirus Disease) [1], that does not allow people to leave their homes, as a result it would be very difficult for those players to practice at home without a coach.

* Corresponding Author. Tel.: +20-100-555-0477.
E-mail address: Bassel1600604@miuegypt.edu.eg

10.1016/j.procs.2020.07.024

(a) Gedan-Barai (b) Age-Uke

Fig. 1: Right And Wrong Movements

The primary goal of this approach is a system that the players, instructor, judges and sporting clubs can use to help them in learning and teaching Karate Kata. The system captures the players' movements in real-time, analyses those movements and generates an evaluation report for the users that informs them how to develop their performance or warn them if they performed a move incorrectly. One of the difficulties in this approach was that the players may execute the movements at different speeds [13]. Another challenge is that the player records a stream of movements that needs to be segmented before it is sent to the classifier. As shown in FIGURE: 1, the difference is clear between the correct move and the wrong move in the two movements that the player had performed.

(a) Hidari gedan-barai (b) Migi chudan oi-zuki (c) Migi gedan-barai (d) Migi tetsui-uchi (e) Hidari chudan oi-zuki (f) Hidari gedan-barai (g) Migi jodan age-uke

Fig. 2: First Seven Movements Of Karate Kata 1 (Hein Shodan) [2]

As shown in FIGURE: 2, the paper is focusing on the first 7 movements of Karate Kata 1 (Hein Shodan), which are arranged as follows: Hidari gedan-barai, Migi chudan oi-zuki, Migi gedan-barai, Migi tetsui-uchi, Hidari chudan oi-zuki, Hidari gedan-barai, Migi jodan age-uke.

The primary contribution of this paper is segmenting the stream of movements, eliminating any noise from the data, interpolating the data, flattening the data to 1D and classifying the movement to identify whether the player performed the move correctly or not. The rest of the paper is arranged as follows: SECTION: 2 explains the researches and approaches of other systems and researchers. SECTION: 3 explains the current approach and which technologies were used. SECTION: 4 shows the examination and evaluation that this approach has reached. Finally SECTION: 5 is the conclusion of the approach and also where the future work were discussed.

2. Related Work

2.1. Hardware And Classifiers

The first related work which supported us in choosing which IR camera sensor to use was written by T. Hachaj et al. [8] and [11] they wrote a paper on how to use the Kinect to determine and identify the Karate movements with the best performance. Initially, they stated the differences between the two versions of the Kinect in some movements. They stated clearly that the Kinect v2 would maintain better results in the identification of the movements and is more reliable. Moreover, they proposed a method for collecting the skeleton data using three Kinects instead of one, so that the movement is captured from all the angles and combined into one skeleton. They encountered a problem which was increasing the movement recognition accuracy of the data collected from the three Kinects. They discovered a position for each Kinect that boosted the non-classified movements by 48%. They conducted various experiments on six data-sets collected from the two versions of Kinects using differently several classifiers in gesture recognition. They acquired the highest accuracy using the Multilayer Perceptron algorithm, but the classification time was 65.93 seconds. Moreover, Bassel Emad et al. [6] wrote a paper on using the Kinect and the F-DTW to detect, analyze and classify Karate Kata individual movements. They were evaluating their approach on four movements. An accuracy of 90% was reached in their approach. However, they noticed that the accuracy decreases in the leg movements, due to the shortage of distance between the leg and the Kinect.

2.2. Movements Evaluation

P. Alborno et al. [3] and N. T. Thanh et al. [19] proposed a method for rating and evaluating the movements based on the quality and execution of each movement. They proposed a system for calculating Karate movements quality. The examination and evaluation of the movements' qualities presents a significant issue encountering people developing smart coaching systems. They proposed a performance scoring system using the Kinect's data. Their goal was to design a model that can be used in any system around the globe for the martial arts.

2.3. Data Enhancement

T. Hachaj et al. in [13] designed an evaluation and visualization approach for advanced gesture analysis. Tests were conducted to evaluate and visualize the similarities and differences between different 3D movements in Karate. Their concern was in inspecting the differences in Karate movements, due to the imperfect imitation issue. E. Escobedo-Cardenas and G. Camara-Chave in [7] using the Kinect, developed a method to recognise the hand gestures. They designed this method to overcome the issue of hand gesture recognition using Sensors, and video based tools. T. Hachaj et al. in [9] designed a method that detects the behavior descriptions with maximally three key-frames. Making the motion detection in low-dimensional feature space remain their primary goal, and how to select the dominant features to distinguish which human behavior it is. Y. Choubik et al. in [5] designed a real-time gestures classification technique. Their goal was to apply machine learning algorithms to the real-time recognition of human body poses. T. Hachaj et al. in [10] developed a technique that enables both numerical and categorical features computation. Their goal was to design an effective application for Karate learning.

3. Methodology

This paper is a continuation and improvement of our previous work [6]. As shown in FIGURE: 3, the system consists of a Kinect connected to a computer that has an internet connection. The player starts by logging in to his account, then calibrates his positioning to the Kinect, so that the joints could be captured correctly. After everything is set-up and the performer is prepared to perform the moves, the performer will click on the start recording button so that the Kinect could start recording his movements by capturing the frames then extracting the skeleton data from these frames to be saved in a CSV file. After the performer finishes, the performer will click on the stop recording button. Then the segmentation and the pre-processing phases will occur on the CSV file so it can be forwarded to the processing stage. The processing will classify the movement using the stream analysis. The stream analysis is where the segmented movements that the player has performed is classified. After the processing is done, a class will be assigned to each movement, then the results are sent to be stored in the database. The results are also displayed briefly on the computer's screen. It also can be shown in detail in a report which contains all the performer's information and also all the details of the movements that the performer has done, such as the movement's name, movement's duration, movement's score, their performance so far. The detailed report could be personalized to show what the user really needs to know.

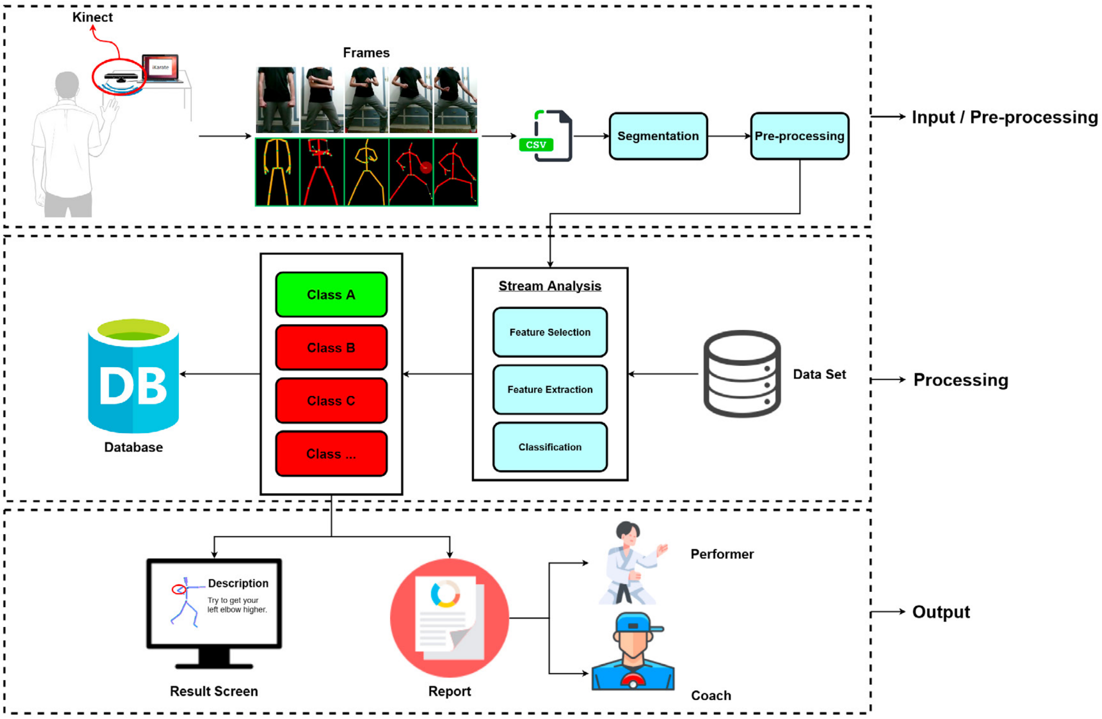


Fig. 3: System Overview

3.1. *Input And Pre-processing*

3.1.1. *Input*

The input of the system is a CSV file that contain the movements data, which is the 25 joints of the skeleton detected by the Kinect in 3D coordinates (X, Y, Z) as mentioned in SECTION: 3.4. Each dimension of each joint is stored in a column. After extracting the data from the CSV file the pre-processing phase is done. The primary phase in the pre-processing is the stream segmentation phase.

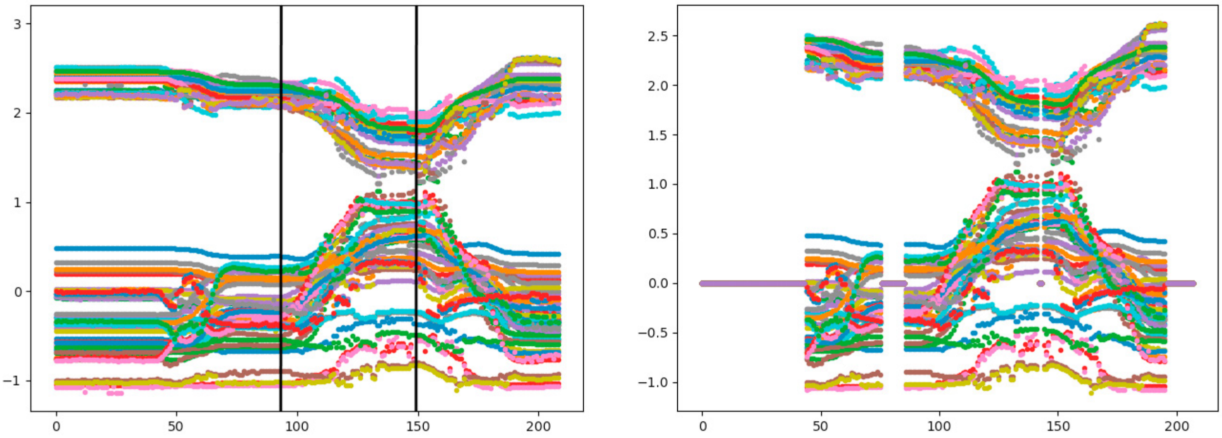
3.1.2. *Segmentation*

The segmentation is the part of the system that will separate the stream of movements of the player to individual movements to obtain an exact accuracy of his performance on each movement. Karate Kata consists of sequences of movements, and each movement should be analyzed independently and separated from the other movements to accurately process each one of them without interfering with each other.

As the player performs each move in Karate Kata there is a slight pause before transitioning to the next move, that was concluded after conducting some experiments and observing the data. Hence, a segmentation method was made according to the information collected. The approach implemented to segment the stream of movements was to create a Queue that will act as a window filled with joint coordinates and then calculate the mean of each two consecutive windows and subtract them and calculate their absolute value as shown in EQUATION: 1, if they are below the threshold it will indicate that the player is not performing any kind of motion.

$$Difference = |Mean(Window_1) - Mean(Window_2)| \tag{1}$$

The Threshold is calculated automatically using **ISODATA** algorithm which is an unsupervised clustering method that takes all the differences between the windows and will detect the value at which the player is steady.



(a) Before Segmentation

(b) After Segmentation

Fig. 4: Segmentation Results

As shown in FIGURE: 4, the left part: 4a shows a representation of our ground truth data without being segmented, and the two vertical lines represent a flag between each movement that was inserted manually for experiment purpose. To the right part: 4b demonstrates how our segmentation method successfully partitioned each movement.

3.1.3. Noise removal and Interpolation

After doing the segmentation and getting the windows that points to the transaction parts between each movement, only the first two frames of the window were removed to avoid corrupting the movement signal. After that there was unwanted data in the middle. So any signal that is less than 30 frames which referees to 1 second will be ignored. Interpolation was used for adjusting the data before building the SVM, K-NN and the DT models, so linear interpolation was used to make all movements equal in size and maintain the signal shape. The features in our data were the joints coordinates which will be flattened to 1D array for the classifier to accept it.

3.2. Processing And Algorithms

3.2.1. Fast Dynamic Time Warping

F-DTW is an algorithm for comparing differences between two signals [17]. Each movement is likely to be performed at different speeds which makes every signal different from the other in size, so an algorithm to handle this problem was necessary. F-DTW is used to handle the player's various performing speeds captured by the Kinect and to present the player with an optimal real-time evaluation. It is also used in [13], [7], [9], [14], [4], [12], [15] and [6]. F-DTW can be extended to various data forms, such as images, audio and graphics. Therefore, any signal that can be transformed to a linear sequence could be analyzed using it.

After extracting the data from the Kinect and pre-process this data, as mentioned in SUBSECTION 3.1.1, F-DTW takes all the data in the file which is composed of 75 columns representing the 3D coordinate of each joint.

$$D(P) = \sum_{K=1}^{\kappa=K} D(P_{ki} - P_{kj}) \tag{2}$$

F-DTW computes the distance and path as shown in EQUATION: 2. Where **D(P)** is Euclidean Distance of the warp path **P**. Where the wrap path length is **K**, and **D(pki, pkj)** represents the distance of two data points (one from Training data and the other is from the Testing data) in the **kth** element of the warp path as mentioned by Stan Salvador et al. in [17]. F-DTW can compute the distance with multiple methods, but the **Euclidean Distance** is the one this approach is using.

3.2.2. Support Vector Machine

SVM (Support Vector Machine) is a supervised learning classifier. It is effective in recognizing patterns and can be applied to recognize these patterns in a various type of applications. It creates a hyperplane that represents the maximum margin that splits the examples into classes [20], [18].

The data inserted to the SVM are the 3D coordinates of each joint represented in 75 columns, but pre-processing was required before building the SVM model as mentioned in the SECTION: 3.1.3.

3.2.3. K Nearest Neighbours

K-NN (K Nearest Neighbours) is a supervised learning algorithm. The algorithm was trained using the labelled data provided to the algorithm. Given a labelled data-set that is made from observations (X, Y), the algorithm captures the relationship between X which is the data and Y which is the label. The data was pre-processed as mentioned in the SECTION: 3.1.3, so that it could be sent to the classifier to build the model. K-NN has many ways to calculate the distance between the classes, but the Manhattan distance was the one used in the current approach.

3.2.4. Decision Tree

DT (Decision Tree) is a supervised algorithm that is used in the data mining and creating classification systems. Which is based on classifying data into segments to construct a tree that consist of nodes and leaf nodes (Classes). The data needed pre-processing before constructing the classifier, and the pre-processing was the same as in the SVM.

3.3. Output

After the system has processed the data and determined the class of each movement, the results are stored in a MySQL database so that it can be viewed later on. The results are also placed in a detailed report so that the player and coach can view it and see the performance of the player.

3.4. Kinect Camera Sensor

The equipment used in this methodology is the Kinect IR camera sensor. The software used in utilizing and using it was the Microsoft SDK (Software Development Kit) as stated in [21] and [16]. Using the benefits of the Infrared Emitter to capture the performer's movements and extract the skeleton easily, also the 3D coordinates of 25 Joints.

Additionally, as mentioned in Microsoft [16], the Kinect is able to deliver 30 frames per second with a 640 x 480-pixel. It is capable of measuring the depth using the depth camera sensor which can tell the performers how far away they are from the camera to calibrate their position. It starts by capturing the person's RGB (Red Green Blue) colors to produce the shape only. Then to acquire the third dimension of the player's skeleton, the monochrome sensor and the infrared projector receive the rays emitted.

4. Experiments And Results

In order to train the classifiers, a data-set had to be collected first. The data-set is collected by two 21 and 24 years old black-belt professional players. It contains the first seven moves of Karate Kata 1 (Hein Shodan). The collected data overall was 210 trials, 105 trials for the correct movements and another 105 trails for a common mistake in the movements. Accordingly, each move has 30 trials, 15 trails for the correct performing and the other 15 are for the wrong performing.

16 trails were made for the testing data, each one represents the sequence of the seven movements done in the correct order which gives us 112 movements. Each testing trial was a combination between the correct performing and

the wrong performing. The system was evaluated with all the classifiers mentioned in the processing SECTION: 3.2, to show the best fitted supervised classification algorithm with the data-set and the segmentation method mentioned in the segmentation SECTION: 3.1.2 and as shown below in the FIGURE: 5, F-DTW has reached the highest accuracy of 91.07% while SVM has an accuracy of 81.25% further, the K-NN has reached 73.21% and finally the DT got an accuracy of 63.39%.

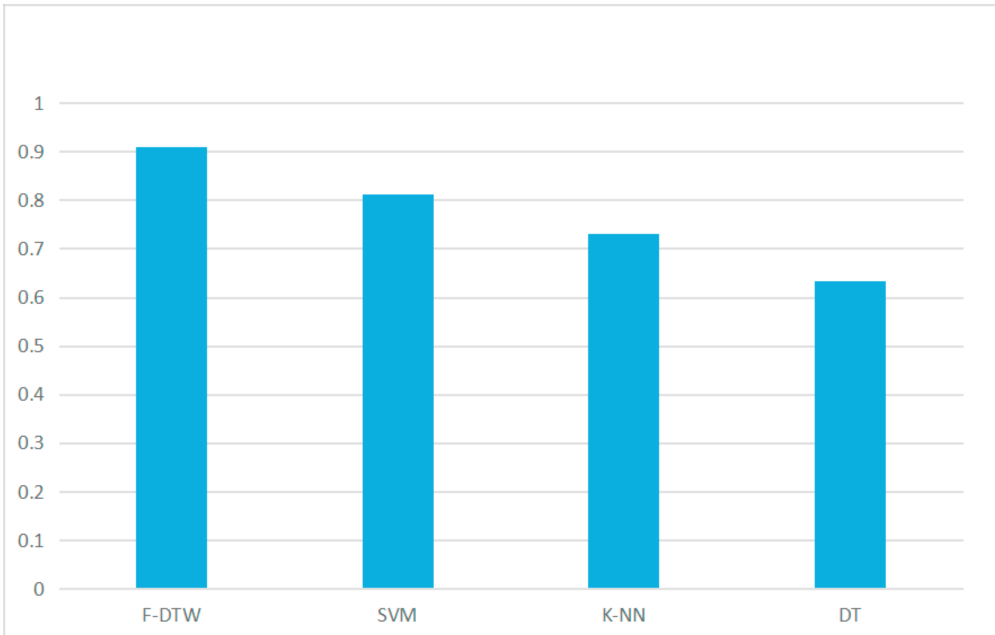


Fig. 5: Overall Accuracies

Table 1: F-DTW Movements Results

Move	Trials	Correct	Wrong	%
Hidari gedan-barai	16	15	1	93.75%
Migi chudan oi-zuki	16	14	2	87.5%
Migi gedan-barai	16	15	1	93.75%
Migi tetsui-uchi	16	15	1	93.75%
Hidari chudan oi-zuki	16	15	1	93.75%
Hidari gedan-barai	16	15	1	93.75%
Migi jodan age-uke	16	13	3	81.25%

TABLE: 1 demonstrates the accuracy of each movement. The average accuracy calculated from the F-DTW as it was the best classifier is 91.07%. This accuracy has been discussed with a Karate coach to determine whether is it good enough or not. The coach confirms that this accuracy is very good implying the Karate Kata system.

5. Conclusion And Future Work

From the related work it has been established that the Kinect sensor is effective at acquiring the skeleton of the player and the joints in 3D coordinates (X, Y, Z). The achievement done in the pre-processing was the segmentation of the first seven movements in Karate Kata 1 (Hein Shodan), taken from two experienced performers. Those moves are partitioned into individual movements, then F-DTW is applied to classify the movements of the whole stream. F-DTW was capable of detecting each movement with an overall accuracy of 91.07%. Finally, showing the performance and a report to the performer. As for future work, the system could implement a deep learning algorithm for better results. Moreover, the system could use socket programming to deliver the data to be classified instead of the CSV file for real-time results.

Acknowledgments

This work was supported by Al-Ahly sporting club and its performers, we thank all the performers that assisted in this approach.

References

[1] , . Coronavirus. https://www.who.int/emergencies/diseases/novel-coronavirus-2019, last accessed on 15/04/20.
[2] , . Shotokan karaté do. https://shotokancrsa.com/heianshodan-eng.htm, last accessed on 16/04/20.
[3] Alborno, P., Giorgis, N.D., Camurri, A., Puppo, E., 2017. Limbs synchronisation as a measure of movement quality in karate. Proceedings of the 4th International Conference on Movement Computing - MOCO 17 doi:10.1145/3077981.3078027.
[4] Calin, A.D., 2016. Variation of pose and gesture recognition accuracy using two kinect versions. 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) doi:10.1109/inista.2016.7571858.
[5] Choubik, Y., Mahmoudi, A., 2016. Machine learning for real time poses classification using kinect skeleton data. 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV) doi:10.1109/cgiv.2016.66.
[6] Emad, B., Atef, O., Shams, Y., El-Kerdany, A., Shorim, N., Nabil, A., Atia, A., 2020. ikarate: Improving karate kata. Procedia Computer Science 170, 466–473. doi:10.1016/j.procs.2020.03.090.
[7] Escobedo-Cardenas, E., Camara-Chavez, G., 2015. A robust gesture recognition using hand local data and skeleton trajectory. 2015 IEEE International Conference on Image Processing (ICIP) doi:10.1109/icip.2015.7350998.
[8] Hachaj, T., Ogiela, M.R., Koptyra, K., 2015a. Effectiveness comparison of kinect and kinect 2 for recognition of oyama karate techniques. 2015 18th International Conference on Network-Based Information Systems doi:10.1109/nbis.2015.51.
[9] Hachaj, T., Ogiela, M.R., Koptyra, K., 2015b. Human actions modelling and recognition in low-dimensional feature space. 2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA) doi:10.1109/bwcca.2015.15.
[10] Hachaj, T., Ogiela, M.R., Koptyra, K., 2015c. Learning from annotated video: An initial study based on oyama karate tournament recordings. 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC) doi:10.1109/3pgcic.2015.16.
[11] Hachaj, T., Ogiela, M.R., Piekarczyk, M., 2013. Dependence of kinect sensors number and position on gestures recognition with gesture description language semantic classifier. Proceedings of the 2013 Federated Conference on Computer Science and Information Systems .
[12] Hachaj, T., Ogiela, M.R., Piekarczyk, M., 2014. Real-time recognition of selected karate techniques using gdl approach. Advances in Intelligent Systems and Computing Image Processing and Communications Challenges 5 , 99–106doi:10.1007/978-3-319-01622-1_12.
[13] Hachaj, T., Ogiela, M.R., Piekarczyk, M., Koptyra, K., 2017. Advanced human motion analysis and visualization: Comparison of mawashi-geri kick of two elite karate athletes. 2017 IEEE Symposium Series on Computational Intelligence (SSCI) doi:10.1109/ssci.2017.8285269.
[14] Hachaj, T., Piekarczyk, M., Ogiela, M.R., 2018. How repetitive are karate kicks performed by skilled practitioners? Proceedings of the 2018 10th International Conference on Computer and Automation Engineering - ICCAE 2018 doi:10.1145/3192975.3192989.
[15] Janaki, V.M.S., Babu, S., Sreekanth, S.S., 2013. Real time recognition of 3d gestures in mobile devices. 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS) doi:10.1109/raics.2013.6745463.
[16] Miles, R.S., 2012. Start here!: learn the Kinect API. Microsoft.
[17] Salvador, S., Chan, P., 2004. Fastdtw: Toward accurate dynamic time warping in linear time and space.
[18] SHUJUN HUANG, NIANGUANG CAI, P.P.P.S.N.Y.W.W.X., 2017. Applications of support vector machine (svm) learning in cancer genomics. Cancer Genomics Proteomics doi:10.21873/cgp.20063.
[19] Thanh, N.T., Tuyen, N.D., Dung, L., Cong, P.T., 2017. Implementation of technical data analysis of skeleton extracted from camera kinect in grading movements of vietnamese martial arts. 2017 International Conference on Advanced Technologies for Communications (ATC) doi:10.1109/atc.2017.8167625.
[20] Yue Shihong, Li Ping, H.P., 2003. Svm classification its contents and challenges doi:https://doi.org/10.1007/s11766-003-0059-5.
[21] Zhang, Z., 2012. Microsoft kinect sensor and its effect. IEEE Multimedia 19, 4–10. doi:10.1109/mmul.2012.24.