



Averaging of motion capture recordings for movements' templates generation

Tomasz Hachaj¹ · Katarzyna Koptyra² ·
Marek R. Ogiela²

Received: 24 March 2017 / Revised: 26 April 2018 / Accepted: 9 May 2018 /
Published online: 24 May 2018
© The Author(s) 2018

Abstract In this paper we propose, describe and evaluate the novel motion capture (MoCap) data averaging framework. It incorporates hierarchical kinematic model, angle coordinates' preprocessing methods, that recalculate the original MoCap recording making it applicable for further averaging algorithms, and finally signals averaging processing. We have tested two signal averaging methods namely Kalman Filter (KF) and Dynamic Time Warping barycenter averaging (DBA). The propose methods have been tested on MoCap recordings of elite Karate athlete, multiple champion of Oyama karate knockdown kumite who performed 28 different karate techniques repeated 10 times each. The proposed methods proved to have not only high effectiveness measured with root-mean-square deviation (4.04 ± 5.03 degrees for KF and 5.57 ± 6.27 for DBA) and normalized Dynamic Time Warping distance (0.90 ± 1.58 degrees for KF and 0.93 ± 1.23 for DBA), but also the reconstruction and visualization of those recordings persists all crucial aspects of those complicated actions. The proposed methodology has many important applications in classification, clustering, kinematic analysis and coaching. Our approach generates an averaged full body motion template that can be practically used for example for human actions

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11042-018-6137-8>) contains supplementary material, which is available to authorized users.

✉ Tomasz Hachaj
tomekhachaj@o2.pl

Katarzyna Koptyra
kkoptyra@agh.edu.pl

Marek R. Ogiela
mogiela@agh.edu.pl

¹ Pedagogical University of Krakow, Institute of Computer Science and Computer Methods, 2 Podchorazych Ave, 30-084, Krakow, Poland

² AGH University of Science and Technology, Cryptography and Cognitive Informatics Research Group, 30 Mickiewicza Ave, 30-059, Krakow, Poland

recognition. In order to prove it we have evaluated templates generated by our method in human action classification tasks using DTW classifier. We have made two experiments. In first leave - one - out cross - validation we have obtained 100% correct recognitions. In second experiment when we classified recordings of one person using templates of another recognition rate 94.2% was obtained.

Keywords Signal averaging · Movements' templates · Motion capture · Kalman filter · Dynamic time warping · Barycenter averaging · Karate

1 Introduction

Many researches on physical activities are based on evaluation of a single or several kinetic or kinematic parameters like maximal force, velocity or acceleration for which mean value and standard deviation among participants is calculated [16, 64, 66]. Even when an empirical model is presented, very often it does not apply full body [27, 39, 49, 57, 58, 65]. This is a big simplification of those models because for example in martial arts the elite fighters use whole body to get the high speed and impact of techniques. The aspects like, for instance, movement's trajectory that is a crucial part of accurate technique performance is often not taken into account. There are of course some exceptions from that trend [47], however very often experiments are performed on low precision hardware and it is impossible to get accurate results. What is more, often the evaluation of kinematic models is done on group of sportsmen that have long experience and it is assumed that they perform some action "correctly" and "optimally" [50]. But while analyzing the results, we see shallowness of evaluation: we know the average results on some activities, however we know nothing about the averaged body motion trajectories with which that results were obtained. In this paper we will present the motion capture signal averaging techniques which allow to generate actions' templates from a group motion capture (MoCap) recordings. Those templates have many potential applications, for example kinematic analysis, coaching, actions classifications, clustering etc.

The human motion's models based on training on MoCap data typically use Markov Models [42], graph representation [13] or Dynamic Time Warping (DTW) [1, 51, 61]. Very often those methods do not employ full body evaluation or operate in reduced PCA space [14] in which not all features are taken into account during calculation.

Authors usually use forward kinematic model to determine the position and orientation of the body parts with given joints angles. On the other hand, we can estimate joint angles with desired or known position of body parts when the inverse kinematic model is used [56]. In our paper we decided to use local coordinate systems (hierarchical model) for angles calculation instead of projection of angles on sagittal, frontal and transversal planes [66]. Thanks to this we simplify the movements features calculation, because in hierarchical model data (beside to root joint) is invariant to any outer coordinate system. Of course hierarchical model can be recalculated to kinematic model that uses angles description relative to common fixed axis and vice versa. However if we would like to compare two MoCap recordings of the same action gathered from two persons that perform movements facing different direction (for example when they vector that links shoulders of one person is perpendicular to the same vector designed by shoulders of the other person) we will obtain different angles values. In case of hierarchical kinematic model all angles besides Hips can be directly compared without additional calculation. That is because of the fact that Hips is a root of the hierarchical model and only this joint is responsible for the global rotation

of the whole body. Also the local coordinate system is more intuitive for coaches and athletes. It is easier to explain the movement (especially three-dimensional actions!) relatively to other parts of the body than to virtual planes which are more usable when we describe one- or maximally two-dimensional actions.

We have tested two signal averaging approaches. The first is Kalman Filter (KF) [34], which is very popular signal processing algorithm with many applications, for example in Earth science [43], biomedical engineering [31], robotics [40], kinematic model synthesis [8, 11, 33, 37] or object tracking [60]. The second averaging approach is based on DTW barycenter averaging (DBA) [53], which was already initially applied to movements' analysis [41, 59]. The evaluation of our new approach was done on karate techniques MoCap dataset. There were two main reasons why we have chosen this particular type of physical activities. At first it can be observed that even a few months of sport training can visibly increase the improvement of gait and posture of people that do various activities, like fitness, Tai Chi or Karate [55]. Due to this fact martial arts training becomes more and more popular in all group ages from preschool children to retiree. Because of the large popularity of that sport there is also a need of computer aided methods and applications that support training [67, 69]. The second reason is that karate has a large group of well defined "standardized" attacking and defense techniques that are practiced by athletes. The model technique performance constitutes the natural template of that action. By averaging a group of MoCap recordings of the same actions we can not only compare the averaged performance of an athlete to the model performance, but also numerically generate that template.

The main novelty of this paper is a proposition and evaluation of motion capture data averaging framework. It incorporates hierarchical kinematic model, angle coordinates' preprocessing methods, that recalculate the original MoCap recording making it applicable for further averaging algorithms, and finally signals averaging processing. We have tested two signal averaging methods namely Kalman Filter (KF) and Dynamic Time Warping barycenter averaging (DBA). We have tested our method on MoCap recordings of elite Karate athlete, multiple champion of Oyama karate knockdown kumite who performed 28 different karate techniques repeated 10 times each. The proposed methods proved to have not only high effectiveness measured with root-mean-square deviation and normalized Dynamic Time Warping distance, but also the reconstruction and visualization of those recordings persists all crucial aspects of those complicated actions. The method we introduce in this paper incorporates reliable and already known approaches (KF and DTW); however, if they are applied in accordance with our research idea it results with valuable practical output. Among them are for example classification, clustering, kinematic analysis and coaching. In order to prove this we successfully use templates generated by proposed framework for human action recognition task. In this paper we also introduce the "Last-chance" nonlinear averaging algorithm which is also our contribution. For our best knowledge there were no other published papers that subject were averaging of whole body motion capture recordings of the same activity for movements' templates generation. In order to generate motion capture templates our algorithm requires several high-quality MoCap recordings of one person that performs an actions to be averaged several times. However we did not find available dataset that satisfy those needs to serve as general benchmark for our research. Because of that we decided to make our own dataset and make it available for download [68]. It can be used as the reference dataset for future research.

In the next section we will define the problem we want to solve, dataset we used for methodology validation, MoCap hardware and kinematic model that was used to acquire this dataset. We also introduce data preprocessing methods and template generation algorithms we have designed. The third section presents validation results of our approaches. In fourth section

we discuss the results by analyzing obtained numerical and visual data. The last section summarizes the paper, shows potential applications and directions of further researches.

2 Material and methods

Let us assume that we have two time series:

$$X := (x_1, x_2, \dots, x_N), Y := (y_1, y_2, \dots, y_M); \quad (1)$$

that represent the same human activity (movement) of the same body joint (for example angle of right hand rotation about Z-axis), but acquired in different MoCap recordings. A warping path (N, M) is a sequence $p = (p_1, \dots, p_L)$ with:

$$p_l = (n_l, m_l) \in [1, N] \times [1, M] \quad \text{for } l \in [1, L] \quad (2)$$

that satisfies the following conditions:

1. Boundary condition: $p_1 = (1, 1)$ and $p_L = (N, M)$.
2. Monotonicity condition: $n_1 \leq n_2 \leq \dots \leq n_L$ and $m_1 \leq m_2 \leq \dots \leq m_L$.
3. Step size condition:

$$p_{l+1} - p_l \in (1, 0), (0, 1), (1, 1) \quad \text{for } l \in [1 : L - 1].$$

Intuitively, the sequences are warped in a nonlinear fashion to match each other [48]. However, if $N = M$ and $p_{l+1} - p_l = (1, 1)$, the warping is linear. Our goal is to generate a new time series' set that is averaged from many time series representing the same human activity. We will take into account two cases: first, in which we consider that signals can be wrapped linearly, and second, in which signals will be wrapped nonlinearly. The averaging algorithms will be named linear averaging algorithm and nonlinear averaging algorithm appropriately.

2.1 MoCap hardware and kinematic model

To do a MoCap, we have used Shadow 2.0 wireless motion capture system based on accelerometers. We have chosen sensor-based MoCap system, because in case of optical-based tracking even with a highly professional system there are many instances where crucial markers are occluded or when the algorithm confuses the trajectory of one marker with that of another [20]. The basic configuration of system consists of 17 sensors placed over body on special costume. With the help of third-party applications and our software we generated the hierarchical kinematic model of human body that is visualized in Fig. 1. This model is very similar to one used typically in popular Biovision Hierarchy (BVH) file format.

In hierarchical model we used rotations of body joints are described relatively to their parent joints in tree-like fashion. The in our case (see Fig. 1) the root joint is a hips joint. The lower body hierarchy goes as follows:

Hips \rightarrow Thigh (left or right) \rightarrow Leg (left or right) \rightarrow Foot (left or right).

The upper part of the body is:

Hips (left or right) \rightarrow SpineLow (left or right) \rightarrow SpineMid (left or right) \rightarrow Chest (left or right) \rightarrow Neck (left or right) \rightarrow Head;

and:

Chest (left or right) \rightarrow Shoulder (left or right) \rightarrow Arm (left or right) \rightarrow Forearm (left or right) \rightarrow Hand (left or right).

The hierarchy description is symmetrical for left and right leg and hand.

Finally we did not use data from hands and feet sensors. These signals were omitted because our athlete performed actions bare-handed and barefoot and we had no possibility to attach sensors to those body parts and to keep them in place during kicks and punches done in full speed. The MoCap system we used for data acquisition is based on internal sensors. In this technology it is crucial for tracking sensors to remain in fixed position on human body. It has appeared that our MoCap costume construction does not prevent hands and foot sensors falling off from bare limbs. After falling off, hands and foot sensors were freely dangling, making acquired signals useless. This is of course limitation of our MoCap acquisition hardware. The valid hands and feet tracking data could be processed with our method. The measurements from points named SpineLow, SpineMid and Neck are interpolated from neighbors' sensors by Shadow software. In Fig. 1 we have presented the local coordinates systems of each sensor (all of them are right-handed). In the right-bottom part of the figure we have also presented orientations of rotations in right-handed coordinate system. The tracking frequency was 100 Hz with 0.5 degrees static accuracy and 2 degrees dynamic accuracy. All further angles measurements in this paper will be presented in degrees. The signals counter domain is $[-180, 180)$.

2.2 Dataset

The dataset we used in our experiment are MoCap recordings of elite Karate athlete, multiple champion of Oyama karate knockdown kumite. She performed several karate techniques:

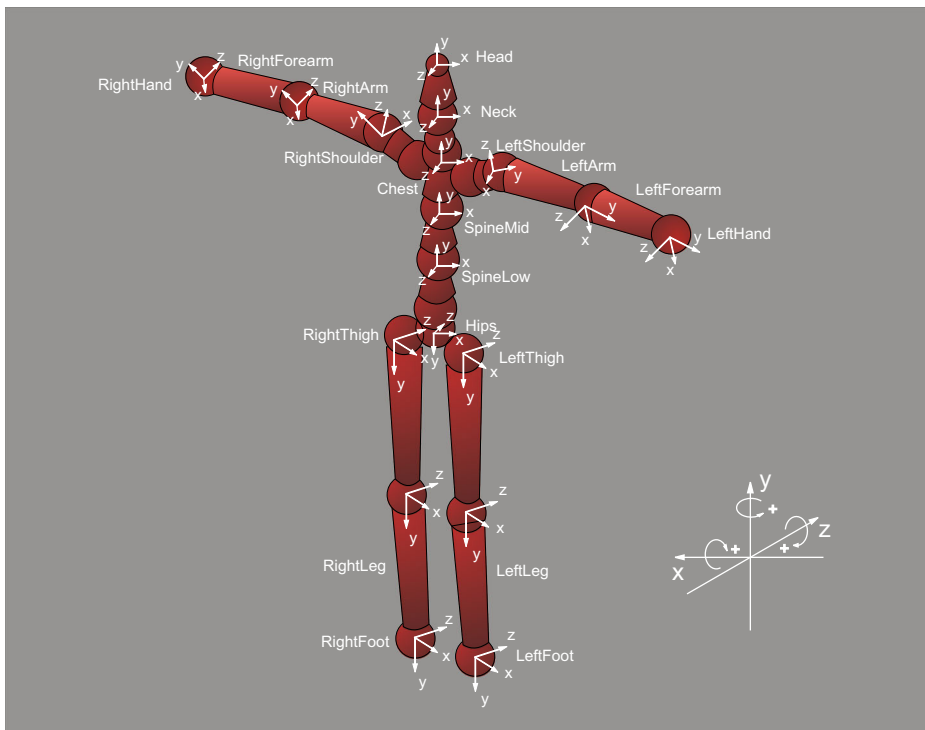


Fig. 1 This figure presents hierarchical kinematic model that we used in our method. The local coordinates systems of sensors are right-handed. In the right-bottom part of the figure we have also presented the orientation of rotation in right-handed coordinate system

- Karate stands: kiba-dachi, zenkutsu-dachi, kokutsu-dachi. The stands were preceded by fudo-dachi. The stands were done in left and right side, so there were 6 types of recordings.
- Defense techniques: gedan-barai, jodan-uke, soto-uke and uchi-uke. Those techniques were done with left and right hand, so there were 8 types of recordings.
- Kicks: hiza-geri, mae-geri, mawashi-geri and yoko-geri. Those techniques were done with left and right leg, so there were 8 types of recordings.
- Punches: furi-uchi, shita-uchi and tsuki. Those techniques were done with left and right hand, so there were 6 types of recordings. In all punches the rear hand was used.

Athlete performed ten repetition of each action. Between each recording the MoCap system was calibrated to maintain the adequate motion tracking. Then the acquired data was segmented into separate recordings in which each sample contains only single repetition of an action.

2.3 Data preprocessing

An algorithm that we will propose in this paper allows to perform Euler angles averaging. Euler angles are commonly used in biomechanics and kinematic analysis [27, 42, 58, 65, 66], because they are very intuitive and easy to interpret rotation description method. In case we are working on time varying signals averaging we have to deal with two factors: signals might vary in length and periodicity. The MoCap signal is commonly saved with counter domain limited to 360 degrees. It is enough for rotation description; however it might be insufficient for instant distance – based comparison of Euler angles. Without this comparison it is impossible to perform signal averaging. We will explain this with the help of images from our dataset.

In Fig. 2 we present plots of hips rotation about a Z axis while performing mawashi-geri kick with right leg. The motion capture frequency was 100 Hz, so of course we have to remember that plots do not represent continuous functions. We decided to represent the acquired data with line plot to increase the visibility and show the trends. The rapid transition from minimal to maximal values are caused by periodicity of angle function which has counter domain $[-180, 180)$, however all plots represents nearly identical rotation. As can be seen, without preprocessing that considers periodicity those signals cannot be aligned with a distance – based method. Our MoCap signals preprocessing is consisted of following steps:

1. signal resampling,
2. angle correction,
3. angle rescaling.

In signal resampling step angles measurements of all recordings in data set are resampled to have the same length. The length equals the number of samples of the longest signal. We use the nearest neighbor interpolation. We have chosen this method because it does not introduce new values to signals. The introduction of a new value (especially by averaging two samples that values are on the border sites of signal counter domain) might eliminate the information that is necessary for the next algorithm that is angle correction.

Angle correction step eliminates the signals discontinuities caused by periodicity of angles functions. The algorithm eliminates large signal's values jump and changes counter domain from $[-180, 180)$ to \mathbb{R} . The pseudo code on Algorithm 1 and Algorithm 2 describes our approach. Algorithm iterates through sampled values of an input signal *signalToCorrect* and checks if changes between two neighboring samples are greater than

a threshold value *tolerance*. If it is so, we have detected case when due the periodicity there is a large difference in angle value and we have to correct it by adding or subtracting by 360 degrees. However; even if MoCap sampling is high (like 100Hz in our case) there is a possibility, that there would be large changes in signals values that happens for example during very fast movements like punches. In this case we cannot apply the correction, because the original signal is already correct. Without visualization of the whole body activity we cannot simply guess (by looking only on rotation data) with which of these two situations we are dealing with. Because of this automatic (black-box) optimization of threshold value is virtually impossible. Due to this tolerance parameter in Algorithms 2 was chosen after basic trial and error evaluation. The results of application this algorithm on signals from Fig. 2 is presented in Fig. 3. As can be seen we have eliminated all discontinuities; however the first signal is in the different period than all the other and we cannot yet apply distance – based signal comparison.

Algorithm 1 This is a helper function, that rescale angle to $[-180, 180)$ domain.

```

Data: angle – an angle to be rescaled
a := angle;
if  $a \geq 180$  then
  | while  $a \geq 180$  do
  | | a := a - 360;
else
  | if  $a < -180$  then
  | | while  $a < -180$  do
  | | | a := a + 360;
return a;

```

Algorithm 2 This algorithm eliminates the signals discontinuities caused by periodicity of angles functions.

```

Data: signalToCorrect – an array with input signal
Result: outS – the output value of the algorithm
outS := signalToCorrect;
tolerance := 120; // when absolute value of signal's value jump
is greater than tolerance the correction is introduced
correction := 0;
for  $a$  in  $2:length(signalToCorrect)$  do
  | prev := rescaleAngle(outS[a-1]);
  | actual := rescaleAngle(outS[a]);
  | if  $prev > tolerance$  and  $actual < -tolerance$  then
  | | correction := correction + 360;
  | else
  | | if  $prev < -tolerance$  and  $actual > tolerance$  then
  | | | correction := correction - 360;
  | outS[a] := outS[a] + correction;
return outS;

```

Algorithm 3 This algorithm performs angle rescaling, which recalculate the angles to the same period.

Data: signal1, signal2 – signal1 is a reference signal while signal2 is a signal to rescale. They has to be the same length.

Result: after angle rescaling signal2 should be in the same period as signal1
tolerance := 300;

```

for a in 1:length(signal1) do
  if signal1[a] - signal2[a] > tolerance then
    | signal2[a] := signal2[a] + 360;
  else
    | if signal1[a] - signal2[a] < - tolerance then
      | | signal2[a] := signal2[a] - 360;
  return signal2;
    
```

The last preprocessing step, angle rescaling, recalculates the angles to the same period. The pseudo code of Algorithm 3 describes our approach. As the reference signal (*signal1*) for rescaling we have arbitrary chosen the first measurement from the group. Algorithm iterates through reference and the second signal (after resampling they have same length)

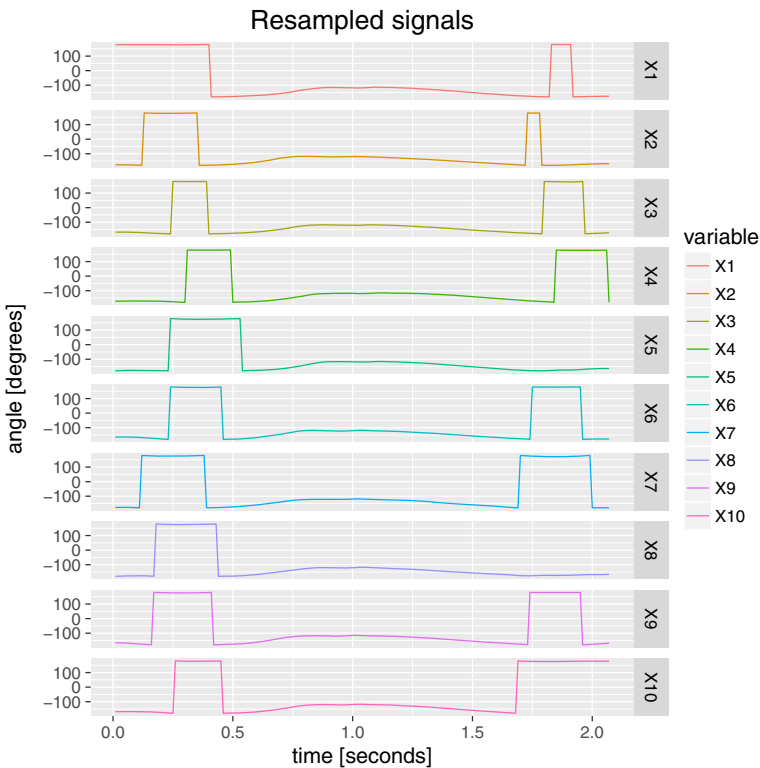


Fig. 2 This figure presents plots of hips rotation about a Z axis while performing mawashi-geri kick with right leg. Each of ten plots came from different MoCap recording. Variables X1, X2 etc. are recordings numbers (there are ten of them). Signals were resampled to have the same length

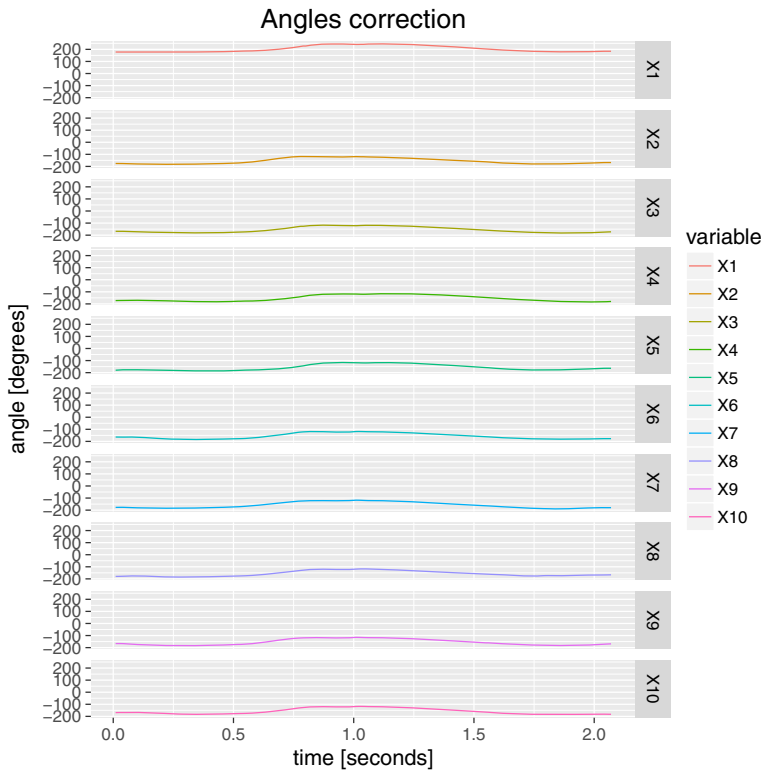


Fig. 3 This figure presents angles correction which should eliminate discontinuities in signal measurements by reducing large signal's values jumps and changing counter domain from $[-180, 180)$ to \mathbb{R} . The presented signals are processed data from Fig. 2

and checks, if the difference between values of corresponding samples is greater than tolerance. If it is so we have to correct second signal by adding or subtracting by 360 degrees. From the same reasons as we explained before we have chosen this parameter after basic trial and error evaluation. Figure 4 presents example results of angle rescaling algorithm on data from Fig. 3. After this step data can be averaged with distance – based methods.

2.4 Linear averaging algorithm

Let us assume that measured MoCap signals from various sensors can be wrapped optimally linearly. In practice, in each group of signals that describes rotation function in time domain of the same action, functions values might deviate between recordings. There might be various natural causes of those deviations – it is nearly impossible that a person repeated the same whole body action identically. Of course the more natural (more trained) the action is, the differences between separate measurements become smaller. There can also be deviations that are caused by inaccuracies of MoCap hardware. In our experiment we assume that the only error that might appear is a random error. Because we are observing the natural phenomena, we will use the state space modeling with the observation from an exponential family, namely Gaussian distribution (dynamic linear models). Let x_t denote the angle values at time t and z_t the set of measured angle positions. Our goal is to estimate x_t using

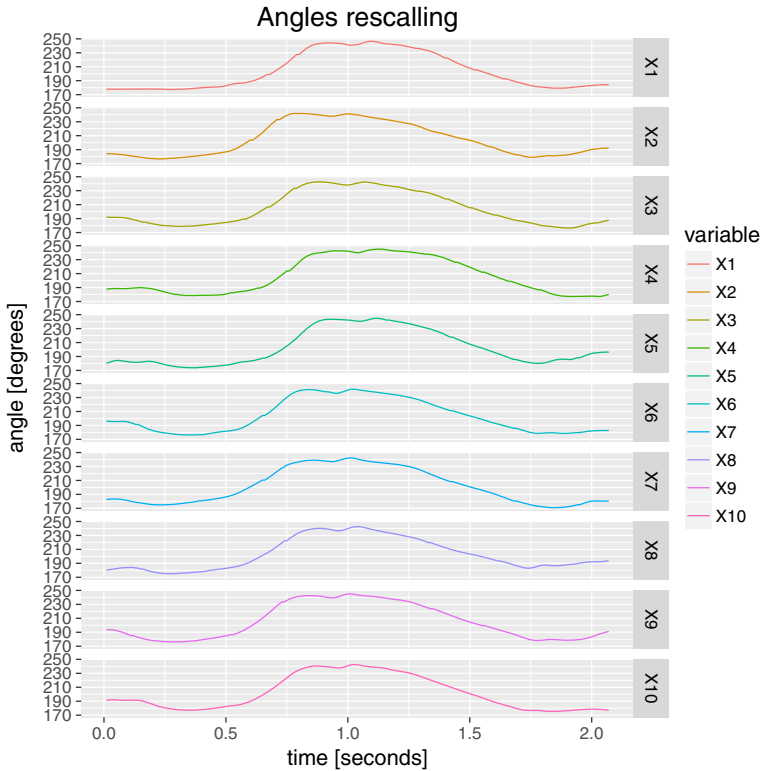


Fig. 4 This figure presents angle rescaling after which angles values should be present in the same period. The presented signals are processed data from Fig. 3

measurements z_t [8, 29]. The Kalman filter is frequently used to solve problems like these, as it provides an optimal solution to tracking problems when states are governed by linear Gaussian motion and observation models. Let us assume that a state x_t evolves as:

$$x_t = F_t x_{t-1} + w_t; \tag{3}$$

with process noise w_t drawn from a zero-mean Gaussian distribution with process covariance Q and a linear transition matrix F_t .

Let us also assume that we can obtain measurements z_t , which are related to the state at time t by the equation:

$$z_t = H_t x_{t-1} + v_t; \tag{4}$$

with measurement of observation noise v_t drawn from a zero-mean Gaussian distribution with measurement covariance R and a linear measurement matrix H_t that maps measurements to states. In our approach we use the smoothing ability of KF to average the signals that came from multiple measurements of the same angle (there are common states for all time series). The unknown parameters of model are calculated with Broyden-Fletcher-Goldfarb-Shanno optimization algorithm in package KFAS [28].

We apply KF to each MoCap signal preprocessed by Algorithms 1–3, and the whole procedure is named linear algorithm (LA). In the situation when signals cannot be wrapped optimally linearly, the nonlinearity between signals will be smoothed by the KF. This

smoothing might visually damage the recording content, because it might treat nonlinearity as noise.

2.5 Nonlinear averaging algorithm

In case when signals can be wrapped optimally nonlinearly, the averaged sequence can be generated by some heuristic method for example DBA [53]. This is a global averaging technique that does not use pairwise averaging. In each iteration DBA performs two steps.

1. Computes DTW between each individual sequence and the temporary average sequence to be refined in order to find associations between coordinates of the average sequence and coordinates of the set of sequences.
2. Updates each coordinate of the average sequence as the barycenter of coordinates associated to it during the first step.

Similarly like in previous linear algorithm in our nonlinear approach DBA is used to calculate each group of signals that describes rotation function in time domain. This time the nonlinear wrapping is not a limitation, however DBA algorithm is a heuristic and we do not have guarantee on quality of the final solution. We will use abbreviation NA (nonlinear algorithm) to this procedure. We have used DBA implementation from R package dtwclust.

2.6 "Last-chance" nonlinear averaging algorithm

Our choice of tolerance parameters in Algorithms 2 and 3 might not fit all possible signal measurements that appear in the dataset. From 28 (techniques count) x 48 (channels count) = 1344 (there are 10 signals among averaging group), seven (about 0.5%) of them were wrongly corrected and due to this wrongly rescaled. An example visualization of this is showed in Fig. 5 where both linear and nonlinear averaging algorithms fail due to failure of angles' correction algorithm. The discontinuity in sample X9 disturbed both algorithms

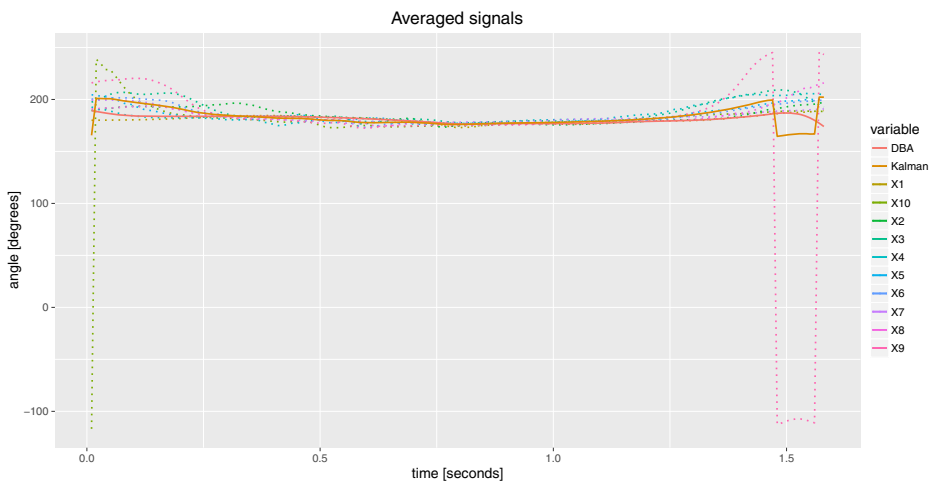


Fig. 5 This figure presents plots of hips rotation about a Z – axis while performing tsuki punch with right hand. Both LA and NA averaging algorithms failed due to failure of angles correction algorithm. The discontinuity in sample X9 disturbed both algorithms based on Kalman and DBA

based on KF and DBA. To deal with this problem without necessity of manual correction of tolerance, we have proposed the "last-chance" angle correction (LCAC, see Algorithm 4) algorithm that base on different principles than Algorithms 2–3.

The LCAC uses information from two channels: one that we want to recalculate and the second one that represents different channel of rotation but of the same joint (sensor). At first sinus and cosines values for both channels are calculated – this step should eliminate the nonlinearity. After this DTW between the first, arbitrary chosen signal from the dataset and the rest of signals from dataset is performed in order to wrap rest of the signals according to DTW results. The DTW is performed for sinus and cosines separately but data is two-dimensional: one dimension is for channel to be corrected and the second channel we mentioned before. This approach increases the stability of final solution by reducing sensitivity of DTW to some local variations of signals to be aligned.

After signal wrapping, KF smoothing is performed on the obtained data. The smoothed data is then filtered with median filter and signal is recalculated from sin/cos counter domain to angle counter domain with inverse trigonometric functions. The example results of the LCAC and its correcting abilities are presented in Fig. 6. It is of course possible that also LCAC fails and we will have to manipulate tolerance parameters of Algorithms 2 and 3, however this situation never happened in our large dataset. We have used DTW implemented [17] with Euclidean distance [15].

Algorithm 4 The "Last-chance" nonlinear averaging algorithm

```

Data: signal1[] and signal2[] – arrays of signals
Result: Alternative to Algorithms 2 – 3.
referenceId := 1;
signal1Sin[] := sin(signal1[]);
signal1Cos[] := cos(signal1[]);
signal2Sin[] := sin(signal2[]);
referenceSignal := [signal1Sin[referenceId], signal2Sin[referenceId]];
for a in 1:length(signal1[]) do
    if a ≠ referenceId then
        querySignal := [signal1Sin[a], signal2Sin[a];
        alignment := DTW(referenceSignal, querySignal); // Dynamic Time
        Warping
        signal1Sin[a] := warp(alignment, signal1Sin[a]); // Signal wrapping
        according to DTW result
        signal2Cos[a] := warp(alignment, signal2Cos[a]);
outSin := MedianFilter(KFSmoothing(signal1Sin[])); // Kalman Filter
smoothing
outCos := MedianFilter(KFSmoothing(signal1Cos[]));
out := asin(outSin);
for a in 1:length(out) do
    if outCos[a] < 0 then
        if out[a] > 0 then
            out[a] := 180 - out[a];
        else
            out[a] := -180 - out[a];
return (out);

```

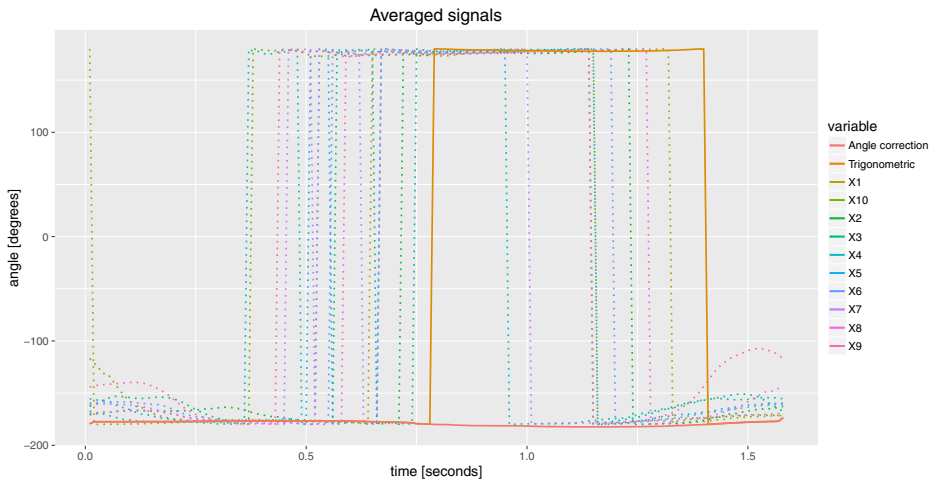


Fig. 6 This figure presents angles' correction which eliminates discontinuities in signal measurements with LCAC. The presented signals is corrected data from Fig. 5

3 Results

We have performed averaging of our dataset described by hierarchical model from Section 2.1 with algorithms described in Sections 2.3, 2.4, 2.5 and 2.6. Example results of this procedure are presented in Fig. 7. The result CSV (comma-separated values) files were then exported to BVH format which can be visualized in many popular rendering programs (for example Blender). The visualizations can be rotated and translated with 6 degrees of freedom which is very helpful in expert evaluation of the results [7, 54].

The proposed approach has been implemented in R language. We have used 'dtw' package [17] that implements DTW and 'dtwclust' for DBA implementation. We have utilized Kalman filter implementation from package 'KFAS' [29]. In DTW we have used Euclidean distance and well-known symmetric2 step pattern. Those are most typical parameters that also in our experiment returned very good results.

The example three-dimensional rendering is presented in Fig. 8 and in supplementary video file [ESM.1](#). That video file presents three-dimensional rendering of averaged mawashi-geri kick with LA and NA. Averaged data is situated in the front of the visualization. In the background of the video there are ten original MoCap recordings that have been averaged. Visualization is playback with 0.25 of the original speed to make movements easier to observe.

3.1 Expert evaluation

The averaged results of our algorithms were evaluated by an expert – a martial artist who indicated errors that she noticed in three-dimensional visualizations. It is possible that an error that has small numerical value might be clearly visible on output rendering. If the expert spotted an error, the LCAC algorithm was used on the particular signal to eliminate the problem. An expert had the following remarks.

- Furi-uchi with the right hand – right hand is positioned over left hand in LA and NA. It has appeared that it is caused by MoCap error so we do not consider it as an error.

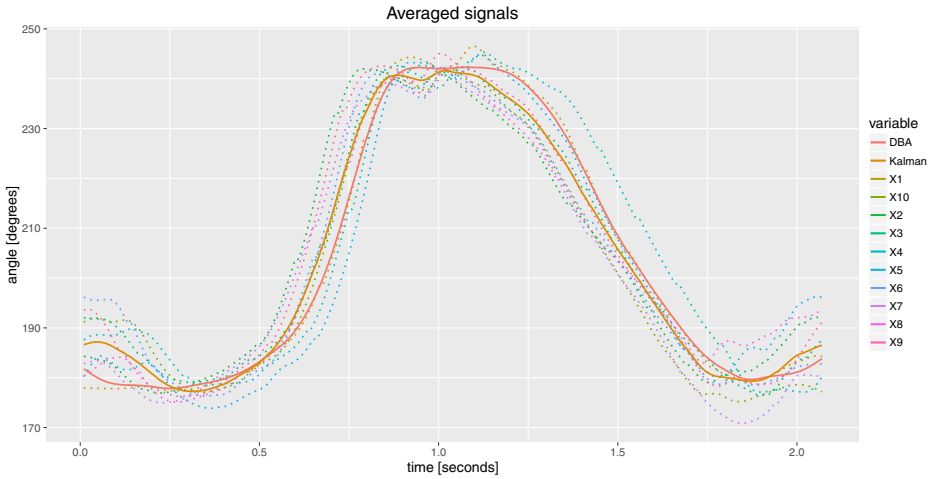


Fig. 7 This figure presents the averaging results with linear LA and NA of signals from Fig. 4

- Gedan-barai with the right hand – the right arm movement is jerky in LA. This is caused by similar situation as described in Fig. 5. The LCAC has corrected this error leaving only small angle’s value jumps at the end.
- Hiza-geri with the left leg – the left arm movement is jerky in LA. This is caused by similar situation as indicated in Fig. 5. The LCAC has corrected this.
- Hiza-geri with the right leg – the left arm movement is jerky in LA and there is a minimal value jump in NA. The whole body posture is shifted to one side however it is caused by MoCap error so we do not consider it as an error.
- Jodan-uke with the left hand – the left arm movement is jerky in LA. This is caused by similar situation as indicated in Fig. 5. The LCAC has corrected this error leaving only small angle’s value jump at the end.
- Tsuki with the right hand – the left arm movement is jerky in LA. There is also error in hips rotation that is depicted in Fig. 5 in LA and NA. The LCAC has corrected this error leaving only small angle’s value jump in left hand LA at the end.
- Yoko-geri with left and right hands – the left and right arms movement is jerky in LA. The LCAC has corrected those errors.

Summarizing, the NA method better overcomes the problems with jerky movements and there are no small signal’s values jumps (discontinuities of trajectories) that appears during

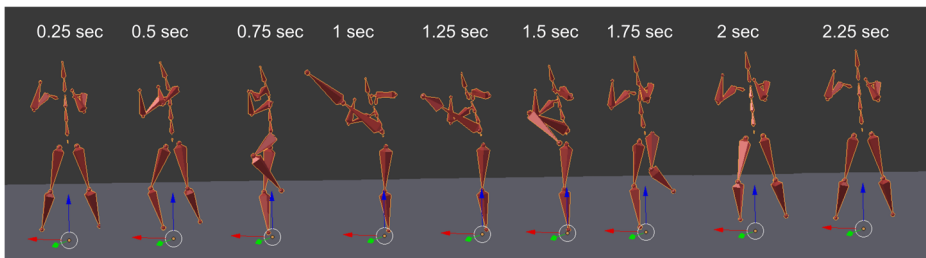


Fig. 8 This figure presents three-dimensional rendering of averaged mawashi-geri kick with LA algorithm

performing averaging of some techniques. The LA algorithm seems to deal with this problem worse, however LCAC eliminates most of the problems sometimes leaving only small signal's values jumps in movement trajectories. Both techniques in expert's evaluation gave good results, but they were not 100% perfect. It seems, however, that NA is more precise in visual reconstruction of original data than LA and expert's recommendation goes in favor to nonlinear solution.

In further numerical evaluation we used the same data as in visual evaluation. When LCAC correction occurred, we have evaluated the improved results.

3.2 Data preprocessing evaluation

Each of 28 karate actions was recorder 10 times and we measured 16 three-dimensional signals (see Section 2.1). Totally we had 4480 signals grouped by 10. We have applied signal preprocessing algorithms from Section 2.3. As we mention in Section 2.6 seven signals among 4480 has failed the angle rescaling step, that means they had different period than others from the same group. The LCAC has corrected those errors. Wrongly calculated signals were the same signals that have been initially spotted by an expert.

3.3 Linear evaluation

This numerical evaluation of our methods is based on comparison of averaged results with original dataset. For each karate technique and for each time sequence we have calculated mean value of root-mean-square deviation (RMSD) between algorithm results and data that was averaged. Thanks to this we could judge how similar averaged signal is to original data. The low value of RMSD also indicates that the optimal signal warping was probably nearly linear.

The root-mean-square deviation formula is:

$$RMSD(X, Y) = \sqrt{\frac{\sum_{i=1}^n \text{diff}(x_i, y_i)^2}{n}}; \quad (5)$$

$$\text{diff}(x_i, y_i) = \begin{cases} \text{abs}(x_i - y_i) & \text{if } \text{abs}(x_i - y_i) < 180 \\ 360 - \text{abs}(x_i - y_i) & \text{if } \text{abs}(x_i - y_i) > 180 \end{cases} \quad (6)$$

The diff formula is the result of the fact, that distance between two angles in degrees is not greater than 180.

3.4 Nonlinear evaluation

For the nonlinear evaluation of averaging we used normalized dynamic time warping distance [48]. Let $c_p(X, Y)$ be a total cost of a warping path p between X and Y with respect to the local measure E (Euclidean distance) defined as:

$$c_p(X, Y) = \sum_{l=1}^L E(x_{nl}, y_{ml}) \quad (7)$$

The optimal warping path between X and Y is a warping path p^* having minimal total cost among all possible warping paths. The DTW distance $DTW(X, Y)$ between X and Y is defined as the total cost of p^* :

$$\begin{aligned} DTW(X, Y) &= c_{p^*}(X, Y) \\ &= \min \{c_p(X, Y) | p \text{ is an } (N, M) - \text{warping path} \} \end{aligned} \quad (8)$$

Normalized DTW is:

$$DTW_n(X, Y) = \frac{DTW(X, Y)}{N + M} \quad (9)$$

Figures 9 and 10 present an example DTW and timeseries alignment from our dataset.

3.5 Application for human actions recognition

Our approach generates an averaged full body motion template that can be practically used for example for human actions recognition. The goal of this type of classification is to assign one predefined class label to recording that contains a person performing particular activity. The proper selection of features that are calculated from motion capture data have a great importance on the final result of classification. There are many papers that propose various possible features sets and compare their performance in pattern recognition tasks. The state-of-the-art discussion on human actions pattern recognition is covered in survey papers for example in [5, 36, 71] or in our previous papers on human actions recognition [25, 26]. In this section we will present overview concerning most popular features sets and pattern recognition algorithms for human motion classification. When the standard video or depth camera data is used for data acquisition, authors often use Gabor filter for features extraction [21], Haar-like features [4] or histogram of oriented gradients and Histograms of Optical Flow [62]. In cases when hardware returns directly spatial coordinates of the body joints authors prefer to use feature that are derived from those coordinates. Among those features are various configurations of angle-based [6, 10, 46] and coordinate-based features [2, 12]. Basing on our previous researches [26] the angle based-features gives better recognition

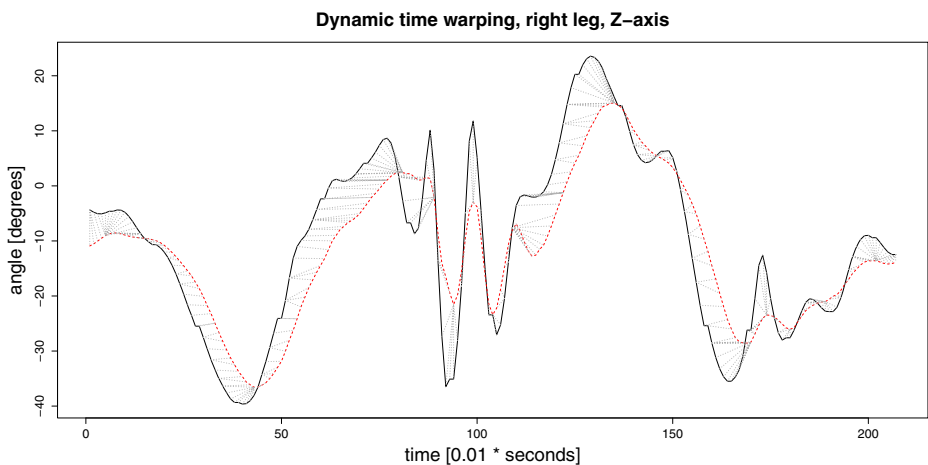


Fig. 9 This figure presents plots of right leg rotation about a Z – axis while performing mawashi – geri kick with right leg. We can see DTW of signal averaged by linear algorithm (red – reference plot) and one of the singles from MoCap recording (query – black plot)

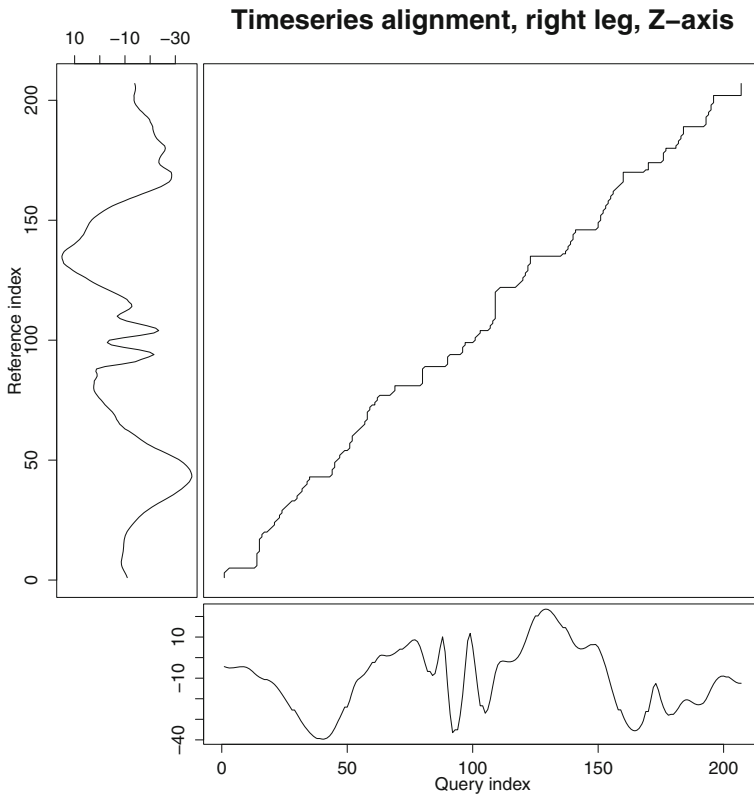


Fig. 10 This figure presents timeseries alignment of data from Fig. 9

results than coordinate - based. The choice of pattern recognition algorithm is often determined by the choice of features set and vice-versa. In case of motion representation with the constant length, in which each action has same number of representation features, no matter how many samples original recording has, authors reports using popular classifiers like for example support vector machines [44], neural networks [32] or even nearest-neighbor [63, 70]. When the continues data stream of MoCap signal classified (like it is in our case) the HMM [26, 45] and DTW [3, 63] are most common approaches. In this paper we have chosen DTW because its application is more straightforward then HMM classifier. That is because DTW does not have additional statistical assumptions, like HMM has, on number of hidden states and distribution (like for example Gaussian) of observable values generated by those states (Table 1).

Table 1 This table summarizes results from proposed algorithms evaluation

jnn	RMSD linear	RMSD nonlinear	DTW linear	DTW nonlinear
MAX	61.76	65.89	26.49	23.40
MIN	0.23	0.26	0.05	0.04
Mean±SD	4.04 ± 5.03	5.57 ± 6.27	0.90 ± 1.58	0.93 ± 1.23
Median	2.60	3.70	0.59	0.66

In our case data is a set of time varying signal acquired with several inertial measurement units (IMU). Each recording of activity of the same class might differ in number of discrete samples, because same activity might be performed with different speed. While using as precise hardware as ours (see Section 2.2 for details) it is virtually impossible for the person to perform some action twice and get the identical recording in temporal and spatial sense. Due to this the most intuitive approach (as we already mention) to solve this pattern recognition task is to use DTW classifier [30]. This pattern recognition method computes similarities between several signal patterns that represents available classes. The input signal is assigned to the class for which there is the smallest dynamic time warping normalized distance (9) between analyzed signal and a template signal that represents this class. We have chosen the similar features set as we successfully used in our previous researches on human actions recognition [23]. In that paper we showed that it is very important to choose the features set that have enough number of dimension with witch it is possible to distinguish between similar actions. From the other hand the number of dimensions cannot be too high because the classifier will be too constrained and the description will lose its generality. It is good to choose the features set that match exactly the most important joints that take direct part in the movement. It is very difficult to prepare the single features set that can be successfully used to classify various human actions that incorporate full body. In the following part of this paper we will use a feature set designed for recognition of karate kicks. This features set is a group of angles calculated between three vectors \bar{x} , \bar{y} and \bar{z} defined relatively to actual body position and vectors representing directions of limbs taking part in the action (thighs and legs). The names of the joints that are used for features definition are present in Fig. 1. At first we will define the non-normalized \bar{x} , \bar{y} and \bar{z} vectors:

$$\begin{cases} \bar{x} = RightThigh - LeftThigh \\ \bar{y} = Hips - Chest \\ \bar{z} = x \otimes y \end{cases} \tag{10}$$

Where \otimes is a vector cross product operator. The legs and thighs vectors are defined as follow:

$$\begin{cases} \overline{ThighL} = LeftThigh - LeftLeg \\ \overline{ShinL} = LeftFoot - LeftLeg \\ \overline{ThighR} = RightThigh - RightLeg \\ \overline{ShinR} = RightFoot - RightLeg \end{cases} \tag{11}$$

Than we calculate angles on plains designed by vectors form (bb1) and (bb2) using following formulas:

$$\begin{cases} A1 = \angle(\overline{ThighL}, \overline{X}) \\ A2 = \angle(\overline{ThighL}, \overline{Y}) \\ A3 = \angle(\overline{ThighL}, \overline{Z}) \\ A4 = \angle(\overline{ThighR}, \overline{X}) \\ A5 = \angle(\overline{ThighR}, \overline{Y}) \\ A6 = \angle(\overline{ThighR}, \overline{Z}) \\ A7 = \angle(\overline{ShinL}, \overline{X}) \\ A8 = \angle(\overline{ShinL}, \overline{Y}) \\ A9 = \angle(\overline{ShinL}, \overline{Z}) \\ A10 = \angle(\overline{ShinR}, \overline{X}) \\ A11 = \angle(\overline{ShinR}, \overline{Y}) \\ A12 = \angle(\overline{ShinR}, \overline{Z}) \end{cases} \tag{12}$$

12 angles we have chosen to classify karate kicks from our dataset nicely covers major possible leg movements of thigh and shin (in our MoCap hardware setup we cannot track the position of feet). As has been shown later in this section, those description are sufficient to make successful classification of lower body motions.

Our DTW classifier works on MoCap data recalculated according to (12). Each sample of recording is 12 – dimensional vector with angle values. As the distance metric we used Euclidean distance. We have performed two experiments: in first we used mae geri, mawashi geri, hiza geri and yoko geri kicks with left and right leg from dataset described in Section 2.2. We have made leave – one – out cross – validation of the classifier. That means that 9 from 10 available recordings of each class were used to generate the template using averaging algorithm proposed in this paper and 1 from 10 was classified with DTW classifier. This procedure was repeated for each one of ten recordings of each class. In this experiment finally all recordings was successfully assigned to correct class.

In second experiment we investigated if it is possible to use averaged recordings of one person to classify recordings of actions performed by another person. In order to do so we have used another dataset containing karate MoCap. We have used additional dataset of Shorin – Ryu karate master that was acquired with the same hardware as we described in Section 2.2. This dataset can be downloaded from the same location as our primary dataset [68]. Shorin – Ryu dataset contains 10 recordings of a person performing mae geri, mawashi geri and hiza geri kick with left and right leg. We have generated templates of those kicks and use them to classify corresponding six type of recordings form Oyama dataset. Than we used six templates from Oyama dataset to classify recordings from Shorin – Ryu dataset. Results of this evaluation are present in Table 2. The recognition rate of DTW classifier using templates generated by our method was 94.2% what is high value for the complexity level and high similarity of those actions. Six hiza geri kick (knee strike) were misclassified as mae geri (forward kick) actions because the initial knee trajectory of both actions are nearly identical.

4 Discussion

We have to remember that, although recordings that were averaged represented same body actions, it is virtually impossible that they were aligned ideally (we have discussed it in Section 2.4). That means we cannot expect that (5) and (9) will be zero. As can be seen in Table 1 mean value is comparable to standard deviation. This relatively large variance among the kinematic parameter among recordings in our dataset is not a surprise. It is of

Table 2 This table presents the classification results of karate kicks set with DTW classifier. Rows stand for true conditions while columns for actual class assignment. All together there are 20 exemplars of each class (10 for Oyama and 10 for Shorin – Ryu master)

	Mae geri right	Mae geri left	Mawashi geri right	Mawashi geri left	Hiza geri right	Hiza geri left
Mae geri right	20					
Mae geri left		19				1
Mawashi geri right			20			
Mawashi geri left				20		
Hiza geri right	3				17	
Hiza geri left		3				17

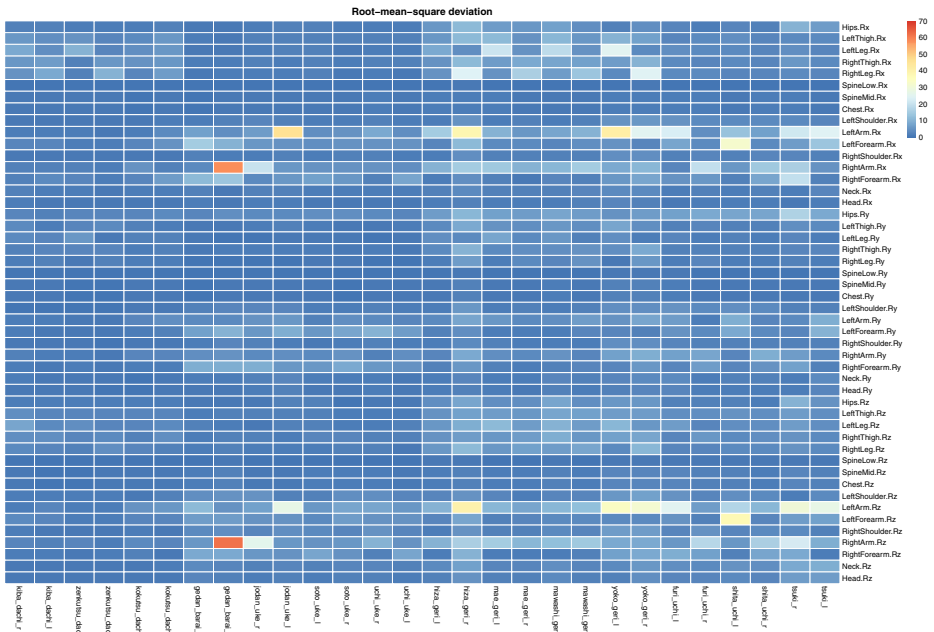


Fig. 11 This figure presents the color-coded plot of mean values of root-mean-square deviation values between signals averaged by LA and all signals from MoCap recordings. In rows there are values’ names and in columns we have techniques’ names

course few times smaller than deviation of kinematic variables in karate group that performs the same activity [65] but it is noticeable even when data come from single, experienced athlete. This is however inevitable attribute of precise MoCap measurement that makes signal processing of full body motion capture data a challenging task. Due to it is hardly possible to mutually align source recordings in averaging process with (5) and (9) below certain thresholds.

The distribution of $RMSD$ and DTW_n can be clearly observed as color coded regions in Figs. 11, 12, 13 and 14. We decided to present results this way because it is much easier to find body actions and body joints that are source of the highest and lowest deviations from average than if the data was tabularized. Each row in those figures represents appropriate measurements of mean value of difference between averaged and input data. Annotations in rows are joints’ names (see Fig. 1). Because rotations are represented by three-dimensional vectors, annotations are grouped by the axis around which rotation is performed. Annotations in columns are names of 28 karate techniques we have in our dataset (see Section 2.2). A result from Table 1 assures us that most of techniques were accurately averaged. Most of the considered cases have relatively low values $RMSD$ and DTW_n . As the result of it we obtained satisfying averaging results and smooth visualizations like one in [ESM.1](#). Now we will discuss the worst case which is gedan-barai defense technique with right hand. In this case in all Figs. 11–14 the highest value is present in right arm joint. Movement of each joint is a result of averaging ten source signals. It is difficult to create an easy to interpreted single plot that incorporate warping results of all time series (similar to one on Fig. 9). Due to this we will discuss this case with the help of supplementary video file [ESM.2](#). This visualization is composed in the same way as [ESM.1](#). The source of high values of (5) and

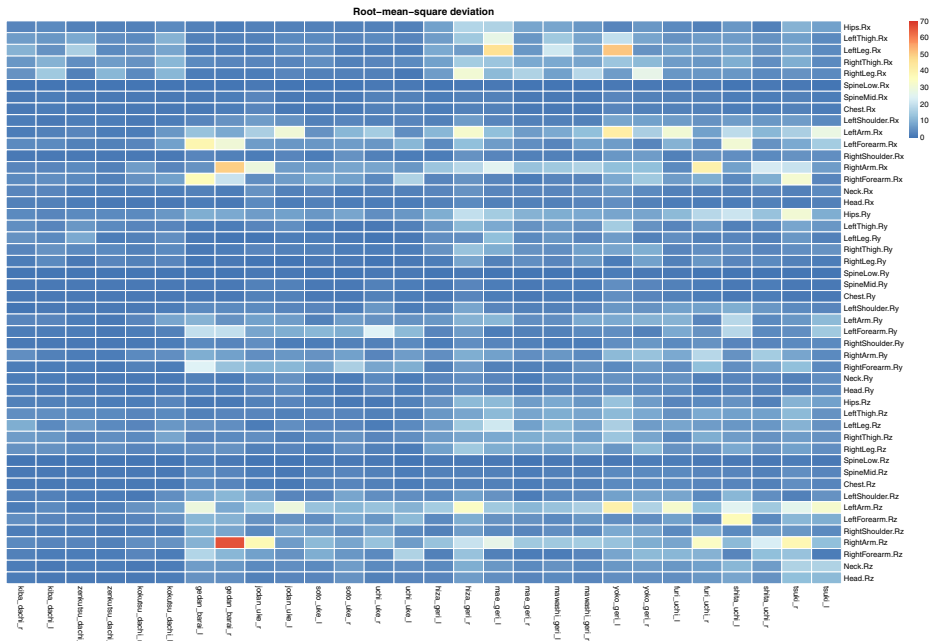


Fig. 12 This figure presents the color-coded plot of mean values of root-mean-square deviation values between signals averaged by NA and all signals from MoCap recordings. In rows there are values names' and in columns we have techniques' names

(9) in case of the LA averaging is clearly visible in the beginning of the second second of the recording, when right arm joint expands to the final position very rapidly. In the same time right arm in NA averaged MoCap changes its position smoothly. The cause of this

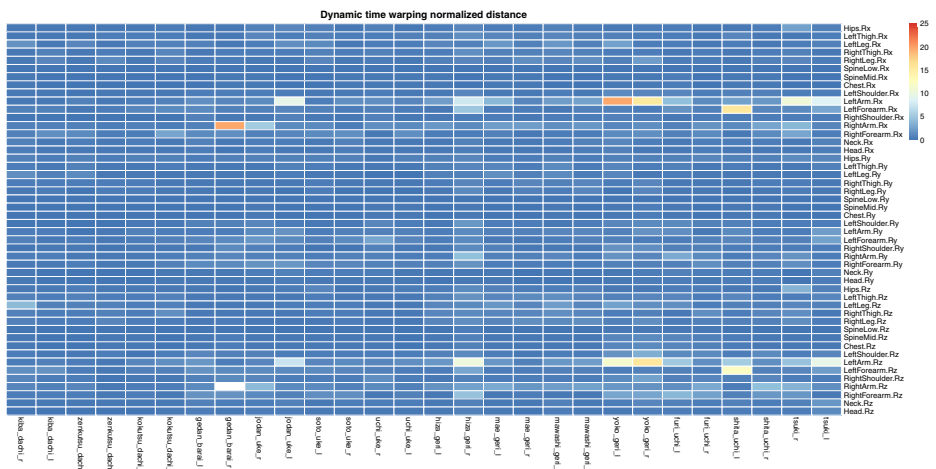


Fig. 13 This figure presents the color-coded plot of mean values of normalized dynamic time warping distance between signals averaged by LA and all signals from MoCap recordings. In rows there are values' names and in columns we have techniques' names

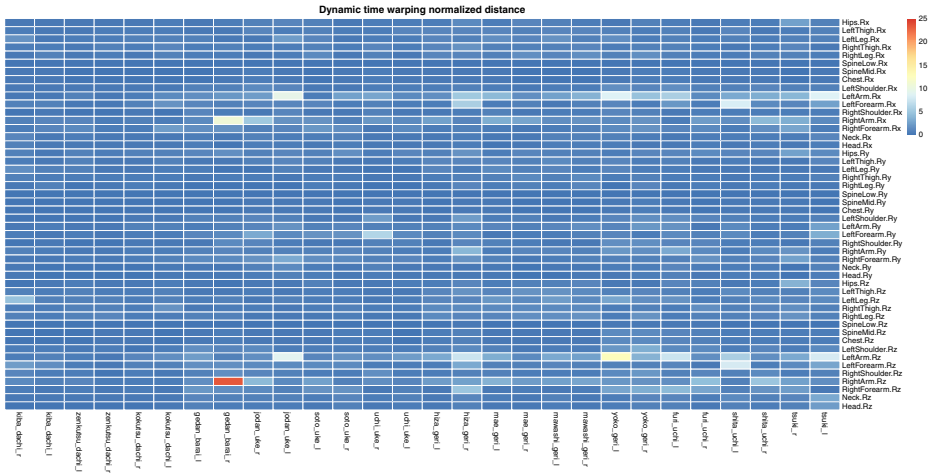


Fig. 14 This figure presents the color-coded plot of mean values of normalized dynamic time warping distance between signal averaged by NA and all signals from MoCap recordings. In rows there are values' names and in columns we have techniques' names

situation is explained when we observe the source MoCap data, that is presented in rear row. The arm movement in most left MoCap visualization ends much later than all others; however this is not a longest MoCap recording that was present in the dataset. This is a situation we mentioned in Section 2.4 – those signals could not be wrapped optimally linearly and KF smoothing damages the recording content, because it treats nonlinearity as noise. In the same time NA have generated similar value for both (5) and (9), however the arm movement is smooth. That is because there is a relatively large variability of data within the source dataset and signal averaging cannot overcome the highest differences. However NA can wrap signal nonlinearly which effect in results that are correct from the perspective of observer (our expert). The very similar situations happen in other cases, when the deviations between averaged data and input MoCaps were relatively high.

As can be seen in Table 1 and Figs. 11–14 both averaging methods resulted in very satisfying numerical results. The mean value of RMSD between averaging results and original MoCap data was 4.04 ± 5.03 degrees for LA and 5.57 ± 6.27 degrees for NA. The median values of those parameters were lower: 2.60 and 3.70 degrees appropriately. It can be explained by the fact that larger mean value together with high standard deviation (larger than mean value) is caused by limited group of techniques that had high movements dynamic in arms (rotation about X and Z axis) and legs (rotation about X and Z axis). Those were gedan-barai, jodan-uke, hiza-geri and yoko-geri. The highest values of deviations in those features can be easily explained. The shoulder and hip joints have anatomical ball and socket construction which allows them to have wide range of rotation about all three axis. In punches and blocks arm position changes rapidly, which in combination with angles' periodicity makes that joints averaging most difficult.

The similar conclusions can be drawn from analysis of DTW normalized distance. The mean value was 0.90 ± 1.58 degrees for LA and for 0.93 ± 1.23 degrees for NA. This time also LA had smaller mean value, however higher standard deviation than NA. It seems that in our case, when the elite karate athlete is analyzed, there are not many nonlinear translocations of actions components in analyzed techniques. Summing up as there are not

much differences in numerical evaluation between LA and NA. Finally we can recommend using the second one, which has better results in expert evaluation.

5 Conclusion

Basing on evaluation in previous sections we can conclude that the proposed methods of MoCap data averaging are reliable and give satisfying results both in visual and numerical evaluation. The proposed methods have many important applications:

- It can be used for template creation for pattern recognition purposes, especially for distance-based methods and clustering [18, 19, 22, 24, 35, 52]. Classification can be done for example with DTW with features set designed in similar way as we presented in Section 3.5 of this paper.
- The averaging ability confirmed by numerical evaluation together with possibility of creating correct visual output make our methodologies usable for coaching purposes. With them a trainer can visualize averaged performance of an athlete and evaluate his or her kinematic parameters [9, 38].
- Many up-to-date researches evaluate only statistical kinematic parameters of actions like mean displacement, velocity or acceleration and their angular analogs. With our approach it is possible to compare and display kinematic differences between various participants of experiments using well established DTW framework (see Figs. 8 and 9).
- With the proposed methodology it is possible to compare the kinematic parameters of actions that were generated in some period of time for example to examine the growth of flexibility during performing dynamic actions. The proposed averaging approach might be a component of evaluation procedure of athlete progress during training.
- The method we proposed is a universal approach which can be applied directly to any MoCap dataset (not only karate) that can be represented with hierarchical model. Methods from Sections 2.3–2.6 can be used without any further generalization to any hierarchical kinematic.

There are also a number of potential applications that have to be examined in further researches. One of the most interesting is application of our methods to compare averaged movements of different athletes in order to compare kinematic parameters of their actions. It might be usable especially in master-student relation, when an adept learns for example karate techniques from a sensei. In that approach athlete tries to make his or her techniques as much similar as it is possible to master's templates movements. After applying computer – aided training we have to evaluate the RMSD and DTW differences between techniques performed by the same person and between various elite martial artists to see what has to be improved. That type of research will give us a knowledge what ranges of parameters one should expect in such evaluation. It might also be useful to generate multi-person templates of popular karate (and other sports) techniques and evaluate numerically and visually those averaged recordings. That research will supply us with interesting and valuable information about kinematic of elite athletes and can be a reference for trainers and scientist. We believe that both proposed method, but especially NA will be an excellent approach to achieve all of those goals.

Acknowledgment This work has been supported by the National Science Center, Poland, under project number 2015/17/D/ST6/04051.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Adistambha K, Ritz C, Burnett IS (2008) Motion classification using dynamic time warping, international workshop on multimedia signal processing, MMSP 2008, October 8–10, 2008, Shangri-la Hotel, Cairns, Queensland Australia, pp 622–627. <https://doi.org/10.1109/MMSP.2008.4665151>
2. Arici T, Celebi S, Aydin AS, Temiz TT (2014) Robust gesture recognition using feature pre-processing and weighted dynamic time warping. *Multimedia Tools Appl* 72(3):3045–3062
3. Arici T, Celebi S, Aydin AS, Temiz TT (2014) Robust gesture recognition using feature pre-processing and weighted dynamic time warping, vol 72
4. Arulkarthick VJ, Sangeetha D (2012) Sign language recognition using k-means clustered haar-like features and a stochastic context free grammar. *Eur J Sci* 78:74–84
5. Berretti S, Daoudi M, Turaga P, Basu A (2018) Representation, Analysis, and Recognition of 3D Humans: A Survey. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 1s, Article 16 (March 2018), 36 pages. <https://doi.org/10.1145/3182179>
6. Bianco S, Tisato F (2013) Karate moves recognition from skeletal motion. In: *Proceedings of the SPIE 8650, Three-Dimensional Image Processing (3DIP) and Applications* Burlingame, CA, USA
7. Bielecka M, Piórkowski A (2015) Automatized fuzzy evaluation of CT scan heart slices for creating 3D/4D heart model. *Appl Soft Comput* 30:179–189. <https://doi.org/10.1016/j.asoc.2015.01.023>
8. Burke M, Lasenby J (2016) Estimating missing marker positions using low dimensional Kalman smoothing. *J Biomech* 49:1854–1858
9. Burns A-M, Kulpa R, Durny A, Spanlang B, Slater M, Multon F (2011) Using virtual humans and computer animations to learn complex motor skills: a case study in karate SKILLS
10. Celiktutan O, Akgul CB, Wolf C, Sankur B (2013) Graph-based analysis of physical exercise actions. In: *Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare (MIIRH '13)*, Barcelona, Catalunya, Spain, 21–25, pp. 23–32 pp
11. ChangWhan S, SoonKi J, Kwangyun W Synthesis of human motion using kalman filter, modelling and motion capture techniques for virtual environments: International Workshop, CAPTECH'98 Geneva, Switzerland, November 26–27, 1998 *Proceedings*, Springer Berlin Heidelberg, pp 100–112, 1998. https://doi.org/10.1007/3-540-49384-0_8
12. Chen X, Koskela M (2015) Skeleton-based action recognition with extreme learning machines. *Neurocomputing* 149(Part A):387–396
13. Endres F, Hess J, Burgard W (2012) Graph-based action models for human motion classification. 7th German Conference on Robotics; *Proceedings of ROBOTIK 2012*, pp 1–6
14. Firouzmanesh A, Cheng I, Basu A (2011) Perceptually Guided Fast Compression of 3-D Motion Capture Data. *IEEE Trans Multimedia* 13(4):829–834. <https://doi.org/10.1109/TMM.2011.2129497>
15. Furlanello C, Merler S, Jurman G (2006) Combining feature selection and DTW for time-varying functional genomics. *IEEE Trans Signal Process* 54(6):2436–2443. <https://doi.org/10.1109/TSP.2006.873715>
16. Gheller RG, Dal Pupo J, Ache-Dias J, Detanico D, Padulo J, dos Santos SG (2015) Effect of different knee starting angles on intersegmental coordination and performance in vertical jumps. *Hum Mov Sci* 42:71–80
17. Giorgino T (2009) Computing and visualizing dynamic time warping alignments in R: The dtw package. *J Stat Softw* 31(7):1–24. <https://doi.org/10.18637/jss.v031.i07>
18. Głowacz A (2015) Recognition of acoustic signals of synchronous motors with the use of MoFS and selected classifiers. *Measurement Sci Rev* 15(4):167–175. <https://doi.org/10.1515/msr-2015-0024>
19. Głowacz A, Głowacz Z (2016) Recognition of images of finger skin with application of histogram, image filtration and K-NN classifier. *Biocybernetics Biomed Eng* 36(1):95–101. <https://doi.org/10.1016/j.bbe.2015.12.005>
20. Guodong L, Leonard M (2006) Estimation of missing markers in human motion capture. *Vis Comput* 22(9):721–728. <https://doi.org/10.1007/s00371-006-0080-9>
21. Gupta S, Jaafar J, Fatimah W, Ahmad W (2012) Static hand gesture recognition using local Gabor filter. *Proc Eng* 41:827–832
22. Hachaj T, Ogiela MR (2012) Semantic Description and Recognition of Human Body Poses and Movement Sequences with Gesture Description Language, *Computer Applications for Bio-technology*,

- Multimedia, and Ubiquitous City: International Conferences MulGraB, BSBT and IUrc 2012 Held as Part of the Future Generation Information Technology Conference, FGIT 2012, Gangneung, Korea, December 16-19, 2012. Proceedings, Springer Berlin Heidelberg, pp 1-8. https://doi.org/10.1007/978-3-642-35521-9_1
23. Hachaj T, Ogiela MR, Koptyra K Human actions modeling and recognition in low-dimensional feature space. In: Proceedings of the BWCCA 2015, 10th International Conference on Broadband and Wireless Computing, Communication and Applications, Krakow, Poland, 4–6, November 2015, pp 247–254
 24. Hachaj T, Ogiela MR, Koptyra K (2015) Application of assistive computer vision methods to oyama karate techniques recognition. *Symmetry* 7(4):1670–1698. <https://doi.org/10.3390/sym7041670>
 25. Hachaja T, Ogiela MR (2015) Full body movements recognition – unsupervised learning approach with heuristic R-GDL method, *Digital Signal Processing*, Volume 46, pp. 239-252. <https://doi.org/10.1016/j.dsp.2015.07.004>
 26. Hachaja T, Ogiela MR (2016) Human actions recognition on multimedia hardware using angle-based and coordinate-based features and multivariate continuous hidden Markov model classifier. *Multimedia Tools Appl* 75(23):16265–16285. <https://doi.org/10.1007/s11042-015-2928-3>
 27. Hadizadeh M, Amri S, Mohafez H, Roohi SA, Mokhtar AH (2016) Gait analysis of national athletes after anterior cruciate ligament reconstruction following three stages of rehabilitation program: Symmetrical perspective. *Gait Posture* 48:152–158
 28. Helske J KFAS: Kalman Filter and Smoother for Exponential Family State Space Models, 2016, <http://cran.r-project.org/package=KFAS> (access date 16 October 2016)
 29. Helske J (2016) KFAS: Exponential family state space models in R, Accepted to journal of statistical software
 30. Huu PC, Le QK, Le TH (2014) Human action recognition using dynamic time warping and voting algorithm. *VNU J Sci Comp Sci Com Eng* 30(3):22–30
 31. Izzetoglu M, Chitrapu P, Bunce S, Onaral B (2010) Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *BioMedical Engineering OnLine* 9(1):934–938. <https://doi.org/10.1186/1475-925X-9-16>
 32. Ji S, Xu W, Yang M, Yu K (2013) 3d convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell* 35:221–231
 33. Jin M, Zhao J, Jin J, Yu G, Li W (2014) The adaptive Kalman filter based on fuzzy logic for inertial motion capture system. *Measurement* 49:196–204
 34. Kalman RE (1960) A New Approach to Linear Filtering and Prediction Problems. *Trans ASME J Basic Eng* 82:35–45
 35. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY (2002) An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Analysis Mach Intell Archive* 24(7):881–892. <https://doi.org/10.1109/TPAMI.2002.1017616>
 36. Ke S-R, Le H, Thuc U, Lee Y-J, Hwang J-N, Yoo J-H, Choi K-H (2013) A Review on Video-Based Human Activity Recognition. *Computers* 2(2):88–131. <https://doi.org/10.3390/computers2020088>
 37. Kok M, Hol JD, Schön TB (2014) An optimization-based approach to human body motion capture using inertial sensors, Proceedings of the 19th World Congress, The International Federation of Automatic Control Cape Town, South Africa, August, vol 24-29, 2014, pp 79–85
 38. Kwon DY, Gross M (2005) Combining body sensors and visual sensors for motion training. Proceedings of the 2005 ACM SIGCHI international conference on advances in computer entertainment technology, pp 94–101. <https://doi.org/10.1145/1178477.1178490>
 39. Lachlan PJ, Haff GG, Kelly VG, Beckman EM (2016) Towards a determination of the physiological characteristics distinguishing successful mixed martial arts athletes: A systematic review of combat sport literature. *Sports Medicine* 46(10):1525–1551. <https://doi.org/10.1007/s40279-016-0493-1>
 40. Larouche BP, Zhu ZH (2014) Autonomous robotic capture of non-cooperative target using visual servoing and motion predictive control. *Auton Robot* 37(2):157–167. <https://doi.org/10.1007/s10514-014-9383-2>
 41. Laurent E, Thomas D, Maïke B, Gavin M (2016) <https://doi.org/10.1080/13658816.2015.1081205>. *Int J Geogr Inf Sci* 30(5):835–853
 42. Lehrmann AM, Gehler PV, Nowozin S (2014) Efficient non-linear markov models for human motion, IEEE conference on computer vision and pattern recognition (CVPR)
 43. Li Y-M, Dong Y-F, Lai M (2007) Instantaneous spectrum estimation of earthquake ground motions based on unscented Kalman filter method. *Appl Math Mech* 28(11):1535–1543. <https://doi.org/10.1007/s10483-007-1113-5>
 44. López-Méndez A, Casas JR (2012) Model-based recognition of human actions by trajectory matching in phase spaces. *Image Vis Comput* 30:808–816

45. Mead R., Atrash A., Matorić MJ (2013) Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction. *Int J Soc Robot* 5:367–378
46. Miranda L., Vieira T., Martinez D., Lewiner T., Vieira AW, Campos MFM (2014) On-line gesture recognition from pose kernel learning and decision forests. *Pattern Recognit Lett* 39:65–73
47. Mitsuhashi K, Hashimoto H, Ohyama Y (2014) The curved surface visualization of the expert behavior for skill transfer using microsoft kinect. *ICINCO* 2:550–555. <https://doi.org/10.5220/0005101305500555>
48. Müller M (2007) *Information retrieval for music and motion*, Springer-Verlag New York, Inc. ISBN 3540740473
49. Neto OP, Magini M, Pacheco MTT (2008) Electromyographic and kinematic characteristics of Kung Fu Yau-Man palm strike. *J Electromyogr Kinesiol* 18:1047–1052
50. Neto OP, Silva JH, Marzullo ANAC, Bolander RP, Bir CA (2012) The effect of hand dominance on martial arts strikes. *Hum Mov Sci* 31:824–833
51. Palma C, Salazar A, Vargas F (2016) HMM and DTW for evaluation of therapeutical gestures using kinect. [arXiv:1602.03742](https://arxiv.org/abs/1602.03742)
52. Peng L, Chen L, Wu X, Guo H, Chen G (2016) Hierarchical complex activity representation and recognition using topic model and classifier level fusion *IEEE transactions on biomedical engineering* PF (99). <https://doi.org/10.1109/TBME.2016.2604856>
53. Petitjean F, Ketterlin A, Gançarski P (2011) A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recogn* 44(3):678–693. <https://doi.org/10.1016/j.patcog.2010.09.013>
54. Piórkowski A, Jajęsnička L, Szostek K (2009) *Creating 3D web-based viewing services for DICOM images, computer networks*, volume 39 of the series communications in computer and information science, pp 218–224. https://doi.org/10.1007/978-3-642-02671-3_26
55. Pliske G, Emmmacher P, Weinbeer V, Witte K (2015) Changes in dual-task performance after 5 months of karate and fitness training for older adults to enhance fall prevention. *Aging Clin Exp Res* 28:1–8. [10.1007/s40520-015-0508-z](https://doi.org/10.1007/s40520-015-0508-z)
56. Qi Y, Soh CB, Gunawan E, Low K-S (2014) A Wearable Wireless Ultrasonic Sensor Network for Human Arm Motion Tracking *Engineering in Medicine and Biology Society (EMBC) 2014 36th Annual International Conference of the IEEE*
57. Quinzi F, Camomilla V, Felici F, Di Mario A, Sbriccoli P (2013) Differences in neuromuscular control between impact and no impact roundhouse kick in athletes of different skill levels. *J Electromyogr Kinesiol* 23:140–150
58. Sbriccoli P, Camomilla V, Di Mario A, Quinzi F, Figura F, Felici F (2009) Neuromuscular control adaptations in elite athletes: the case of top level karateka. *Eur J Appl Physiol* 108(6):1269–1280. <https://doi.org/10.1007/s00421-009-1338-5>
59. Seto S, Zhang W, Zhou Y, Müller M (2015) Multivariate time series classification using dynamic time warping template selection for human activity recognition, *IEEE Symposium Series on Computational Intelligence, SSCI 2015*, Cape Town, South Africa, December 7–10, 2015, pp 1399–1406. <https://doi.org/10.1109/SSCI.2015.199>
60. Shimin Y, Hee NJ, Young CJ, Songhwa O (2011) Hierarchical Kalman-particle filter with adaptation to motion changes for object tracking. *Comput Vis Image Underst* 115:885–900
61. Slama R, Wannous H, Daoudi M (2013) 3D Human Video Retrieval: from pose to motion matching, *eurographics workshop on 3D object retrieval*. <https://doi.org/10.2312/3DOR/3DOR13/033-040>
62. Stasinopoulos S, Maragos P (2012) Human action recognition using Histogrammic methods and hidden Markov models for visual martial arts applications. *Inproceedings of the 2012 19th IEEE International Conference on Image Processing (ICIP)*, Orlando, FL, USA, 30 September–3
63. Su C-J, Chiang C-Y, Huang J-Y (2014) Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic. *Appl Soft Comput* 22:652–666
64. Timmi A, Pennestrì E, Valentini PP, Aschieri P (2011) Biomechanical analysis of two variants of the karate reverse punch (gyaku tsuki) based on the evaluation of the body kinetic energy from 3D mocap data *Multibody Dynamics ECCOMAS*
65. Vences Brito AM, Rodrigues Ferreira MA, Cortes N, Fernandes O, Pezarat-Correia P (2011) Kinematic and electromyographic analyses of a karate punch. *J Electromyogr Kinesiol* 21:1023–1029
66. Vieira P, Moreira S, Goethel FM, Gonçalves M (2016) Neuromuscular performance of Bandal Chagui: Comparison of subelite and elite taekwondo athletes. *J Electromyogr Kinesiol* 30:55–65
67. Vignais N, Kulpa R, Brault S, Presse D, Bideau B (2015) Which technology to investigate visual perception in sport Video vs. virtual reality. *Hum Mov Sci* 39:12–26
68. Website of the GDL project that hosts MoCap dataset we used to validate our method <http://gdl.org.pl/> (Access date: 11.11.2017)

69. Witte K, Emmermacher P, Bandow N, Masik S (2012) Usage of virtual reality technology to study reactions in Karate-Kumite. *Int J Sports Sci Eng* 6(1):17–24
70. Yang X, Tian Y (2014) Effective 3D action recognition using EigenJoints. *J Visual Commun Image Represent* 25:2–11
71. Zhang S, Wei Z, Nie J, Huang L, Wang S, Li Z (2017) A review on human activity recognition using vision-based method. *J. Healthcare Eng* 2017(Article ID):3090343. <https://doi.org/10.1155/2017/3090343>



Tomasz Hachaj received MSc degree in computer science from the Tadeusz Kosciuszko Krakow University of Technology, Poland and PhD degree in computer science from AGH University of Science and Technology, Poland in 2006 and 2010, respectively. He is a deputy Head of the Institute of Computer Science at the Pedagogical University in Krakow, Poland. He is an author of more than 50 papers on pattern recognition, signal processing and image analysis. He specializes in human actions recognition, social networks analysis and medical images processing and visualization.



Katarzyna Koptyra holds the MSc in Computer Science and is currently working toward the PhD degree at AGH University of Science and Technology. She is a scientist developing new information security solutions and steganographic techniques. Her main areas of interest are information hiding and data protection including steganography, cryptography and computer systems security.



Marek R. Ogiela Professor of Computer Science, cognitive scientist and cryptographer. Head of Cryptography and Cognitive Informatics Laboratory in Krakow, Poland. He works at AGH University of Science and Technology, and Pedagogical University in Krakow, Poland. He is a member of numerous world scientific associations i.e. IEEE Senior Member, SPIE Senior Member etc. as well as member of the Interdisciplinary Scientific Committee of the Polish Academy of Arts and Sciences. Author of more than 340 scientific international publications on pattern recognition and image understanding, artificial intelligence, cybernetics, cryptography and information theory.