

Comparison of methods for hand gesture recognition based on Dynamic Time Warping algorithm.

Katarzyna Barczewska
Department of Automatic
Control and Biomedical
Engineering, AGH University
of Science and Technology,
Kraków, Poland
Email: kbarczew@agh.edu.pl

Aleksandra Drozd
Department of Measurement
and Electronics, AGH
University of Science and
Technology,
Kraków, Poland
Email: drozd@agh.edu.pl

Abstract—Gesture recognition may find applications in rehabilitation systems, sign language translation or smart environments. The aim of nowadays science is to improve the recognition systems' efficiency but also to allow the user to perform the gesture in a natural way. The article presents different methods (DTW – Dynamic Time Warping, DDTW - Derivative Dynamic Time Warping, PDTW - Piecewise Dynamic Time Warping) based on Dynamic Time Warping algorithm, which is commonly used for hand gesture recognition using small wearable three-axial inertial sensor. Additionally, different approaches to signal definitions and preprocessing are discussed and tested.

To verify which of the methods presented is more accurate in case of gesture recognition, database of 2160 simple gestures was collected, and recognition procedure was implemented. The main goal was to compare the efficiency of each method assuming that each person should perform the movement naturally. Obtained results suggest that the most efficient method for the presented problem was the DDTW. The worst recognition performance was achieved with the PDTW method.

I. INTRODUCTION

THE recent advance of sensor technologies allows engineers to use smaller and smaller devices capturing human motion. These devices are cameras[1], game controllers, such as Microsoft's Kinect [2] or sensors: inertial [3], [5], [6] or built in data gloves[4]. One of the areas of interest is gesture recognition, which may find its application in game interface design, controlling virtual reality, smart environments but also in biomedical science. For example gesture recognition system may be used as a rehabilitation instrument to improve the sensibility of hands for people recovering from physical accidents or cognitive disabilities [3]. Another application might be an assistive translating system for the deaf people, who use sign languages to communicate [1], [4]. Much research has been done on the topic of gesture recognition, and designers of recognition system should always bear in mind that hand gestures are complicated and the way of performing a gesture varies depending on the person. Solutions presented in the literature using inertial sensor systems reach effectiveness of simple gesture recognition from 69% to 96% for general recognition and from 98% to 99% in recognition of one person's set of ges-

tures [3], [5], [6]. Above solutions, however, assume that all gestures examined are strictly defined or they should be performed in one plane. Our research goal is to compare different variants of widely used for gesture recognition Dynamic Time Warping (DTW) algorithm and conclude which method gives the best results in recognition effectiveness taking into consideration aspects such as: defining the distance between two signals and choosing proper signals for analysis and recognition (acceleration or orientation in space). All tests (in contrast to [3], [5], [6]) are done assuming that a person should do the gesture in a natural way, which implies that a gesture can be performed in 3D space, according to one's preferences and physical conditions. Therefore, small wearable device such as three-axial IMU sensor was considered to collect motion data. Authors based on previous research presented in [7], where information about acceleration and Euler angles in 3-dimensional space of a sensor placed on a forefinger was used to classify the gesture. The aim of this research was to compare different variations of Dynamic Time Warping algorithm (DTW) including Piecewise Dynamic Time Warping (PDTW) and Derivative Dynamic Time Warping (DDTW) in gesture recognition. Moreover, some solutions concerning signal preprocessing leading to the recognition effectiveness improvement are presented.

II. MATERIALS

Authors used database collected for tests described in [7]. Data acquisition was performed using 9 DoF inertial sensor, NEC-TOKIN, Motion Sensor MDP-A3U9S, placed on a volunteer's forefinger (Fig. 1). The small size ($20 \times 20 \times 15$ mm) and weight (6g) of the measurement module allowed user to move his hand and to bend the forefinger in natural way. The sensor contains 3-axis accelerometer, magnetometer and gyroscope. The output raw data contain information from each of these 3-axis sensors as well as the angular orientation expressed in Euler angles. Data transfer to the PC was performed via USB wire (sampling rate - 25 Hz). The set of 10 simple gestures was the recognition subject, their scheme is shown on Fig. 2. Similar gestures were also recognized in [3], [5], [6]. The database consisted of 2160 gestures, each in the separate text file, collected during 3 mea-

This work is supported by Ministry of Science and Higher Education in Poland in years 2013-2014



Fig. 1 Measurement module and its attachment to the forefinger. Original picture from [7]

surement sessions. In a single measurement session each of 9 volunteers performed each gesture 8 times.

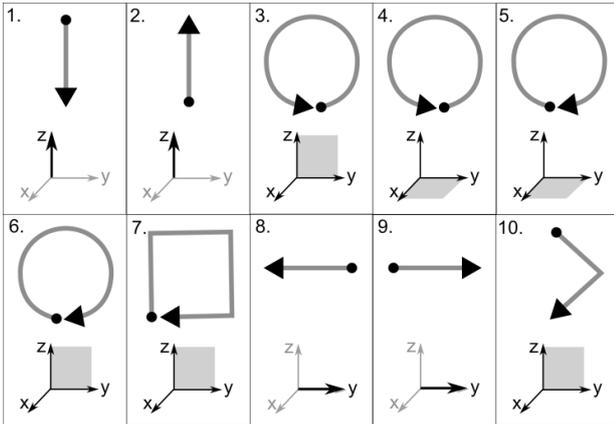


Fig. 2 Schemes of the gestures with coordinate system corresponding to that in measurement module. Main directions and planes of movement were marked.

In view of the fact that the individual gesture performance might vary significantly depending on the time, mood, tiredness or concentration, sessions were held at least one day apart. The database was divided into testing and training sets. 5 of 8 each gesture repetitions performed during measurement sessions were included into the training set, remaining 3 became the testing set. The training set contained 62.5% of all gestures, testing - 37.5%. Contrary to [3] and [6], authors did not assume that during gesture performance palm angular orientation does not change. In gestures made in natural way position changes can be observed, as well as changes of angular orientation. There were no restrictions about the way of gestures performing, what caused that some volunteers performed the same gesture using just one finger, others using the palm and some using whole arm.

III. METHODS

Preprocessing was applied to all measured signals, including mean filtering, signal values scaling to the interval $[-1, 1]$ as well as segmentation to obtain data corresponding only to the activity of performing a gesture uniformed for all examined people. Research on the topic of preprocessing was presented in [7] and showed that signal segmentation based on monitoring changes in Euler angles during movement improves the segmentation process.

Basic algorithm used to both determine exemplar set as well as to classify gestures was DTW algorithm (Dynamic Time Warping). This algorithm is commonly used to find the

similarity between time series. Such a method was chosen because of the characteristic of collected gesture signals which were usually similar but transformed in time. The signals in our data base were different concerning their length, distribution of peak values and the velocity of performing particular gesture phases. Additionally two transformations of the base algorithm were implemented: Derivative Dynamic Time Warping (DDTW) and Piecewise Dynamic Time Warping (PDTW) [7-11].

A. Dynamic Time Warping (DTW)

DTW algorithm allows to compute the distance between two signals in the following procedure. Assuming there are two gesture signals (given by acceleration or angle signals changing in time): $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_m\}$ to align these two sequences using DTW for X and Y we need to define distance matrix D containing Euclidian distances between all pairs of points (x_i, y_j) .

$$D(i, j) = d(x_i, y_j) \quad (1)$$

where

$$d(x_i, y_j) = |x_i - y_j|$$

Then we define cumulative matrix P recursively:

$$P(1, 1) = 0$$

$$P(i, 1) = D(i, 1) + P(i - 1, 1)$$

$$P(1, j) = D(1, j) + P(1, j - 1)$$

For $i, j > 1$

$$P(i, j) = D(i, j) + \min\{P(i - 1, j), P(i, j - 1), P(i - 1, j - 1)\} \quad (2)$$

As a result of the DTW algorithm optimal total distance between X and Y after alignment was obtained, which is denoted by $q_{DTW} = P(n, m)$.

B. Derivative Dynamic Time Warping (DDTW)

DDTW algorithm is a variation of basic DTW algorithm. When the two series may have local differences in the Y-axis, it is useful to take into consideration derivative of the signals instead of the signals themselves. First, we calculate the estimate of derivatives of the signal. Then, as before according to equation (1) we construct an n -by- m matrix D where the (i_{th}, j_{th}) element of the matrix is the distance $d(x_i, y_j)$ between the two points x_i and y_j and finally calculate $q_{DDTW} = P(n, m)$.

C. Piecewise Dynamic Time Warping (PDTW)

PDTW algorithm uses time series transformed to reduced representation. A time series X of length n can be represented by a time series $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$, where $N < n$ and

the i -th element of \bar{X} can be calculated from the following equation:

$$\bar{x}_i = \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \quad (3)$$

This means that data is reduced from n dimensions to N by averaging data in each of N frames. We denote the ratio of the length of the original time series to the length of the reduced representation by the compression rate c .

$$c = \frac{n}{N} \quad (4)$$

If the compression rate is high, it will reduce the time of performing calculations, but it will flatten the signal as well.

D. Different approaches to signal definitions

Since six different signal components (3 acceleration and 3 angle components) might be taken into consideration, same algorithms were used to compare the results of recognition based only on accelerations or only on angles or on both of the values together to show whether information about orientation in space may improve widely used technique based on acceleration analysis. As DTW algorithms compute distance between two signals, it can be used for two one-dimensional signals. However, the signals can be also treated as three-dimensional considering all acceleration or all orientation components. In this case distance matrix D (1) was modified and it consisted of distances between two points in three dimensional space defined by:

$$d(x_i, y_j) = \sqrt{(x_{i1} - y_{j1})^2 + (x_{i2} - y_{j2})^2 + (x_{i3} - y_{j3})^2}$$

$$\text{for } x_i = (x_{i1}, x_{i2}, x_{i3}), y_j = (y_{j1}, y_{j2}, y_{j3})$$

E. Exemplars

Authors proposed two approaches to the exemplars, that were the basis of gesture recognition: user-dependent and user-independent. In user-dependent approach, using samples from the training set, for each person for each of 10 gestures one gesture was indicated as an exemplar. The gesture from the training set became an exemplar, when it was the most similar to others in terms of one of the warping methods (DTW, DDTW and PDTW). For each warping method one individual (user-dependent) exemplar for every gesture was indicated. Then, taking into account all individual exemplars for each gesture, the one that was the most similar to others became general exemplar used in user-independent recognition. There were different sets of exemplars for acceleration, for angles, for all signals, also for 1D distance function and for 3D distance function.

IV. RESULTS

Gestures classification was performed using the testing set, all described below algorithms: DTW, DDTW and PDTW and corresponding sets of exemplars. Firstly, to recognize gestures authors used all acceleration and Euler angles signals at once. Secondly, to determine which parameter is more important for natural gestures recognition, classifica-

tion was carried out for them separately. There were also two different approaches used in single parameter recognition: 1) distance function in algorithms DTW, DDTW and PDTW was computed separately for each component of the parameter (for x , y and z axis in case of acceleration and for pitch, roll and yaw in case of Euler angles – 1D distance); 2) both acceleration and angles were treated as 3-dimensional signals and the distance in DTW algorithms was calculated like in 3-dimensional space giving one value for all acceleration components and one value for all angles (3D distance). Obtained results were also divided into individual and general cases. Individual are average efficiency values calculated for each person using his or her individual exemplar, general are average efficiency values calculated for everybody using the same general exemplar for each person. Efficiency was calculated as the sum of all correctly recognized gestures divided by the sum of all gestures. Efficiency values for each of the described recognition method are shown in Tab. I and Tab. II.

TABLE I.
CLASSIFICATION RESULTS FOR INDIVIDUAL CASE. GIVEN VALUES ARE RECOGNITION EFFICIENCIES.

Basis for classification	Distance function	DTW	DDTW	PDTW
Acceleration	1D	0,922	0,895	0,923
	3D	0,948	0,947	0,940
Angle	1D	0,757	0,861	0,747
	3D	0,789	0,921	0,784
All signals	1D	0,894	0,937	0,895
	Average	0,862	0,912	0,858

TABLE II.
CLASSIFICATION RESULTS FOR GENERAL CASE. GIVEN VALUES ARE RECOGNITION EFFICIENCIES.

Basis for classification	Distance function	DTW	DDTW	PDTW
Acceleration	1D	0,877	0,785	0,872
	3D	0,854	0,836	0,626
Angle	1D	0,584	0,674	0,579
	3D	0,638	0,822	0,626
All signals	1D	0,794	0,828	0,789
	Average	0,749	0,789	0,698

The PDTW method was conducted for three values of compression rate: 2, 3 and 5. All results for the PDTW method shown in tables above correspond to the compression rate equal to 2. The higher the value of this parameter, the lower recognition efficiency for collected database (the PDTW method might give better results for higher sampling rates or unfiltered signals). It can be observed that the highest value for recognition efficiency for individual case was obtained for classification based on 3-dimensional approach to acceleration signals and distance function – all warping methods lead to the value of 0.94 for this parameter. 3-dime-

sional approach to all other signals brought much better results in comparison to the corresponding methods with 1-dimensional distance function. Taking into account individual and general cases, statistic tests were conducted (at the 95% confidence level), indicating significant differences between results obtained by all described DTW methods. Tests revealed no difference only in one case: comparison of DTW and PDTW with compression rate $c = 2$ for general case. The most efficient warping method is the DDTW and the highest efficiency rate is also for acceleration signals. However, analyzing the various gestures separately, it appears that in some cases, the analysis of the angles lead to more accurate classification. Results of each gesture recognition efficiency for different basis of classification for the DDTW method are presented on Fig. 3.

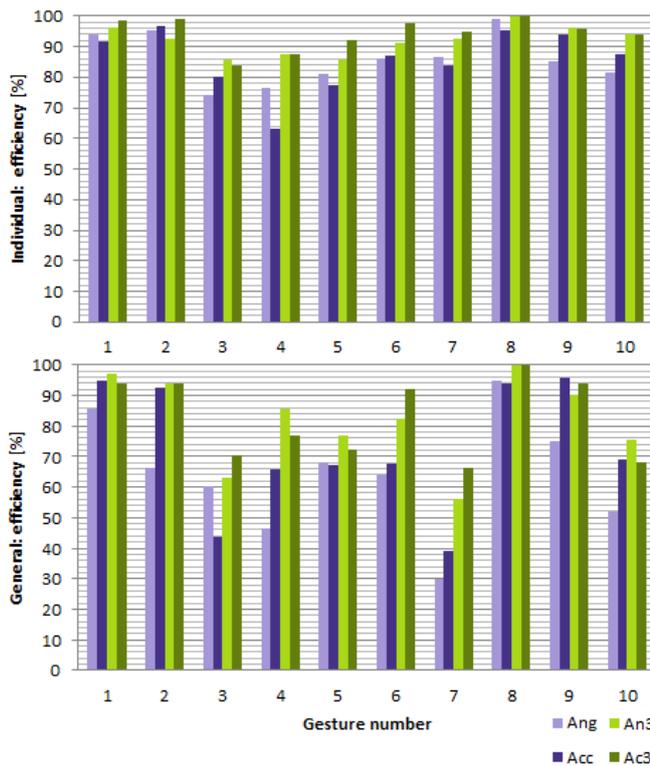


Fig. 3 Each gesture recognition efficiency for individual and general case for different classification basis: Ang – angles, An3 – angles and 3D distance function, Acc – acceleration, Ac3 – acceleration and 3D distance function. DDTW method.

V. DISCUSSION

Authors proved that recognition of gestures performed in a natural way without any constraints for the examined person is possible using methods based on Dynamic Time Warping Algorithms. Analyzing recognition efficiencies obtained for all DTW methods and taking into account the results of statistic tests which revealed significant differences between methods, it can be stated, that the most efficient method to described application is the DDTW. The worst recognition performance was achieved with the PDTW method. The higher compression rate, the higher reduction

of information and the lower efficiency values. The reduction of information that was caused by the PDTW method was a disadvantage in this case, but it can occur that for higher sampling rates it will become an advantage.

Signals collected confirmed the theory that such aspects as time of the day, tiredness, concentration, experience may affect gesture performing which results in lower recognition efficiencies. As these problems are unavoidable, obtaining better results is a matter of algorithms and preprocessing.

Moreover, analyzing Fig. 3 it can be observed that there are gestures which are much better recognized while using angular orientation as a basis of classification. Solution to improve this method to recognize natural gestures would be to combine information about angular orientation and acceleration. Possibly treating the signal as a 6 dimensional (3 acceleration components and 3 angle components) would cause increase of recognition efficiency.

In future, research data base enlarging should be considered to verify the hypothesis which of the methods presented is more accurate in case of general (subject independent) gesture recognition. Another improvement may be creating more general exemplar set containing modeled signals (instead of signals registered by the sensor), which will reduce the influence of between subject variability.

REFERENCES

- [1] Paulraj M. P., Yaacob S., Azalan M. S. Z., Palaniappan R., *A Phoneme Based Sign Language Recognition System using 2D Moment Invariant Interleaving feature and Neural Network*, IEEE Student Conference on Research and Development, 2011.
- [2] Wang Y., Yang C., Wu X., Xu S., Li H., *Kinect Based Dynamic Hand Gesture Recognition Algorithm System*, 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2012
- [3] Hussain S.M.A., Harun-ur Rashid A.B.M., *User Independent Hand Gesture Recognition by Accelerated DTW*, IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision, Proceedings, Dhaka, Bangladesh 2012.
- [4] Liang R.-H., Ouhyoung M., *A Real-time Continuous Gesture Recognition System for Sign Language*, Third IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings, 1998.
- [5] Liu J., Wang Z., Zhong L., Wickramasuriya J., Vasudevan V., *uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications*, Pervasive and Mobile Computing Journal, Vol. 5, Issue 6, Elsevier, Amsterdam, The Netherlands, 2009.
- [6] Akl A., Feng C., Valaee S., *A Novel Accelerometer-Based Gesture Recognition System*, Transactions on Signal Processing, IEEE, vol. 59, No. 12, December 2011.
- [7] Barczewska K., Drozd A., Folwarczny Ł., *Rozpoznawanie gestów z wykorzystaniem czujników inercyjnych o 9 stopniach swobody*, Pomiary Automatyka Kontrola, Vol. 59, No. 3, March 2013.
- [8] Keogh, E., Pazzani, M., *Derivative Dynamic Time Warping*. In First SIAM International Conference on Data Mining (SDM'2001), Chicago, USA
- [9] Keogh, E., Pazzani, M., *Scaling up Dynamic Time Warping for Datamining Applications*. In 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, 2000
- [10] Müller M.: *Information Retrieval for Music and Motion*. Chapter 4: Dynamic Time Warping. Springer Verlag 2007;
- [11] Helwig N. E., Hong S., Hsiao-Weckler T., *Time-Normalization Techniques for Gait Data*, 33rd Annual Meeting of American Society of Biomechanics Materials, State College, PA, USA, 2009.