Motion data classification on the basis of Dynamic Time Warping with a cloud point distance measure

Adam Switonski*, Henryk Josinski†, Hafedh Zghidi† and Konrad Wojciechowski**

*Polish-Japanese Institute of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland
†Silesian Univercity of Technology, ul Akademicka 16, 44-100 Gliwice, Poland
**Polish-Japanese Institute of Information Technology, Aleja Legionow 2, 41-902 Bytom, Poland

Abstract. The paper deals with the problem of classification of model free motion data. The nearest neighbors classifier which is based on comparison performed by Dynamic Time Warping transform with cloud point distance measure is proposed. The classification utilizes both specific gait features reflected by a movements of subsequent skeleton joints and anthropometric data. To validate proposed approach human gait identification challenge problem is taken into consideration. The motion capture database containing data of 30 different humans collected in Human Motion Laboratory of Polish-Japanese Academy of Information Technology is used. The achieved results are satisfactory, the obtained accuracy of human recognition exceeds 90%. What is more, the applied cloud point distance measure does not depend on calibration process of motion capture system which results in reliable validation.

Keywords: motion capture, dynamic time warping, gait identification, classification of motion data, cloud point distance
PACS: 80

INTRODUCTION

In our previous works Dynamic Time Warping (DTW) transform was applied to classification of skeleton based motion data [1, 2]. They are represented by time sequences of relative rotations performed by subsequent joints, coded in a form of Euler angles or unit quaternions. The crucial was the choice of a proper distance function responsible for assessment of dissimilarity between two rotations. The obtained results were satisfactory. For highly precise motion capture measurements and applied validation dataset with 30 classes, over than 90% accuracy is achieved [1]. However in a case of markerless acquisition with much greater noise, it was necessary to consider some additional pose parameters to improve the classification [2]. Distance between right and left ankles and position of head were used. Because of quite different scales and natures of skeleton and non-skeleton pose attributes, the proper choice of coefficients corresponding to their impact on a total distance value is troublesome. It is a reason for investigation of another variant of DTW classification, which takes into consideration model free motion data. They are represented by time sequences of 3D positions of subsequent markers located on human body. In such the case, proper approach to compare cloud points has to be chosen.

The main contribution of the paper is related to the proposed method of motion data classification which utilizes Dynamic Time Warping transform and selected cloud point distance measure. In the validation experiment, collected motion capture gait database containing data of 30 humans is used. What is more, the recognition is also carried out for selected body segments. It allows to assess their discriminative features. Obtained results are compared to the corresponding ones obtained for skeleton motion data, presented in the previous ICNAAM conference [1].

DYNAMIC TIME WARPING WITH CLOUD POINT DISTANCE MEASURE

Dynamic time warping (DTW), introduced long time ago and originally applied to spoken word recognition [3], is very effective method of time series comparison and classification. It outperforms both simple lock-step measures as for instance Euclidean or Manhattan metrics and more sophisticated edit distance approaches - Longest Common SubSequence [4], Edit Sequence on Real Sequence [5], Swale [6], Edit Distance with Real Penalty [7],[8]. Thus, its choice to the problem of motion data analysis seems to be reasonable.

DTW is a elastic measure type, which means that one-to-many mappings are allowed. It matches time instants of compared motion sequences by linear, monotonic transformation to obtain the lowest total cost calculated as aggregation of dissimilarity between matched poses. Thus, DTW takes into consideration possible local shifts between motion phases.

DTW implementation utilizes dynamic programming. In the first stage dissimilarity matrix containing distances between every pair of poses of analyzed motions is calculated. It is used to determine monotonic path with minimal aggregated total distance across underlying pose dissimilarity values, connecting edge points corresponding to start and end of compared time series. Because of monotonicity of the path, the cost to reach the specified poses i and j of the first and second motions respectively can be determined on the basis of costs reaching to possible previous poses with indexes (i-1,j). (i,j-1), and (i-1, j-1). As the result, it is sufficient to carry on the calculations in the proper order for subsequent rows or columns. The final cost of the path found corresponds to motions dissimilarity and the shape of the path to mappings determined.

To calculate similarity matrix, the method to compare time instants has to be established. Aggregated distance d_{naive} between poses P and R for subsequent markers P_i and R_i is unworkable, because it is strictly dependent on the place of acquisition.

$$d_{naive}(P,R) = \sum_i \|P_i - R_i\|^2 \tag{1}$$

Therefore, similar to [9, 10] additional transformation $T_{\alpha,x,z}$, which rotates markers R_i around axis Y and translates them with $(x,0,z)$ vector is taken into consideration. The proper distance d_{CP} is stated to be the minimal sum of squared distances in respect to applied transformation $T_{\alpha,x,z}$:

$$d_{CP}(P,R) = \min_{\alpha,x,z} \sum_i \|P_i - T_{\alpha,x,z}(R_i)\|^2 \tag{2}$$

It means direction of a gait and its location on fixed ground are not used in a proper recognition. The equation (2) can be solved analytically by separating the determination of rotation angle α and offset vector $(x,0,z)$ and finding zeros of derivative function $\frac{\partial d_{CP}}{\partial \alpha}$, which are located in α_1 and α_2 values. (detailed proof is presented in [11]):

$$\alpha_1 = atan\left(\frac{\sum_{i=1}^{K}\frac{1}{K}(x_i^P \cdot z_i^R - z_i^P \cdot x_i^R) - (\overline{x^P}\cdot\overline{z^R} - \overline{z^P}\cdot\overline{x^R})}{\sum_{i=1}^{K}\frac{1}{K}(x_i^P \cdot x_i^R + z_i^P \cdot z_i^R) - (\overline{x^P}\cdot\overline{x^R} + \overline{z^P}\cdot\overline{z^R})}\right) \tag{3}$$

$$\alpha_2 = \pi + atan\left(\frac{\sum_{i=1}^{K}\frac{1}{K}(x_i^P \cdot z_i^R - z_i^P \cdot x_i^R) - (\overline{x^P}\cdot\overline{z^R} - \overline{z^P}\cdot\overline{x^R})}{\sum_{i=1}^{K}\frac{1}{K}(x_i^P \cdot x_i^R + z_i^P \cdot z_i^R) - (\overline{x^P}\cdot\overline{x^R} + \overline{z^P}\cdot\overline{z^R})}\right) \tag{4}$$

where x_i^P, z_i^P, x_i^R, z_i^R correspond to x and z components of marker i of pose P and R respectively and $\overline{x^P} = \frac{1}{K}\sum_{i=1}^{K} x_i^P$, $\overline{z^P} = \frac{1}{K}\sum_{i=1}^{K} z_i^P$, $\overline{x^R} = \frac{1}{K}\sum_{i=1}^{K} x_i^R$, $\overline{z^R} = \frac{1}{K}\sum_{i=1}^{K} z_i^R$ to centroids of markers for P and R poses.

For specified α vector $(x(\alpha),0,z(\alpha))$ has a form:

$$x(\alpha) = \overline{x^P} - \overline{x^R}\cdot cos(\alpha) - \overline{z^R}\cdot sin(\alpha) \tag{5}$$

$$z(\alpha) = \overline{z^P} + \overline{x^R}\cdot sin(\alpha) - \overline{z^R}\cdot cos(\alpha) \tag{6}$$

Finally the transformation $T_{\alpha,x(\alpha),z(\alpha)} \in \{T_{\alpha_1,x(\alpha_1),z(\alpha_1)}, T_{\alpha_2,x(\alpha_2),z(\alpha_2)}\}$ which produces smaller d_{CP} value is chosen.

In Fig. 1 example transformation $T_{\alpha,x(\alpha),z(\alpha)}$ which adjusts randomly selected pose 2 to pose 1 of different subjects is visualized. The rotation and translation are successfully determined - the pose 2 is properly transformed to fit pose 1.

COLLECTED GAIT DATABASE

To validate proposed method motion capture gait database acquired by Human Motion Laboratory of PJIIT `http://hml.pjwstk.edu.pl` equipped with Vicon software and hardware was used. It contains data of 436 gaits of 30 different males. To detect main gait cycle containing two middle adjacent steps, tracking of distances between feet is carried out. The same database was also used in [1]. The default skeleton model of Vicon Blade software was applied as presented in Fig. 2. It consists of 22 joints described by Euler angles triplets, global rotation and three-dimensional

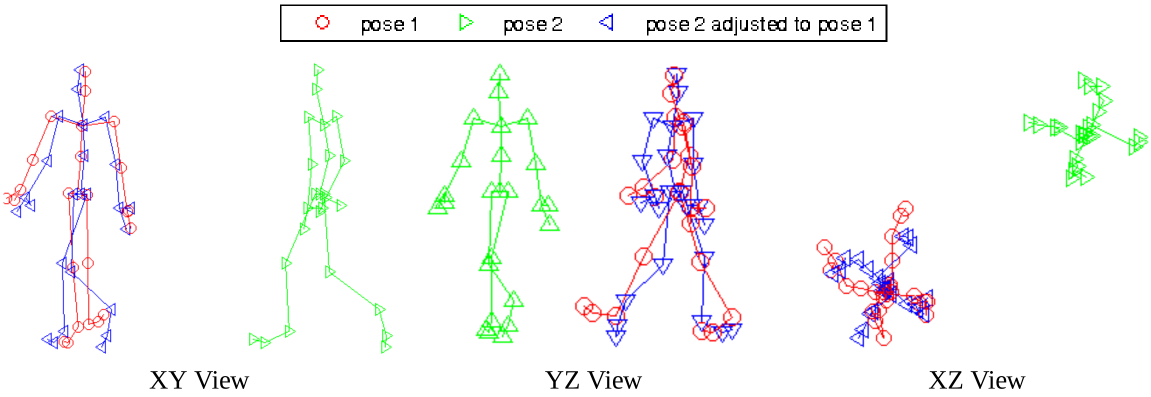

FIGURE 1. Exmaple transformation: pose 2 is adjusted to pose 1

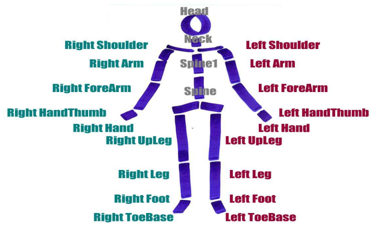

FIGURE 2. Applied skeleton model

translation vector. On the basis of skeleton data, positions of virtual markers are reconstructed. The markers directly utilized in the classification are located on the successive joints of skeleton model as labeled by triangles and circles in Fig. 1. Similar as in [1], to minimize the influence of the calibration process of the motion capture measurements on the final recognition, the acquisition is divided into sessions. The calibration is carried out independently in every session.

RESULTS AND CONCLUSIONS

Utilizing the ability to assess a dissimilarity between two motions by DTW, k nearest neighbor (kNN) classification scheme [12] is directly applied in the recognition. The collected gait database is split into two subsets of approximately equal size - training and testing ones, which contain the data of different sessions. The classification is assessed by CCR ratio - percent of correctly recognized instances of the testing set. In the validation a human gait identification challenge problem is taken into consideration. It means humans are recognized on the basis of their walk characteristics.

The obtained classification results which are compared to the ones achieved with usage of Euler angles and unit quaternions [1] are presented in Table 1. The cloud point distance measure outperforms them for every of considered classifiers - 1NN, 3NN and 5NN. The highest CCR rate is 91%, it means 16 missclassified gait instance of 178. The better results of cloud point distance can be explained by anthropometric features, contained in subsequent positions of the virtual markers. The positions depend not only on the movements of preceding joint, but also segment's length. Thus, such individual traits as human height, the lengths of lower and upper limbs are utilized in the classification. The rotational data give bit less precision, however they describe only the movements of successive joints with much weaker relation to skeleton parameters.

The results of recognition which utilize only some manually selected body segments are shown in Fig. 2. For the sake of compatibility with the outcomes presented in [1], virtual markers are placed at the end of the segment which directly corresponds to movements of specified joint. For instance marker named LeftUpLeg (see Fig. 2) is located in LeftLeg joint position. The selection reduces pose space noticeably, however it also decreases the efficiency of classification. The individual features depend on an activity of the segments during gait, it is consistent with the

TABLE 1. DTW classification on the basis of virtual markers with usage d_{CP} function and rotations of skeleton model.

	Virtual markers	Euler angles [1]		Quaternions [1]	
k	d_{CP}	Euclidean	Manhattan	$d_{geodesic}$	d_{cosine}
1	90,45	86,67	86,67	89,44	83,89
3	91,01	86,11	83,89	87,22	78,89
5	87,64	79,44	81,11	81,67	76,11

TABLE 2. DTW classification for selected segments with d_{CP} distance function.

Markers	CCR
LeftUpLeg;RightUpLeg	84,83
LeftArm;RightArm	80,90
LeftShoulder;RightShoulder	75,28
LeftFoot;RightFoot	66,85
LeftHand;RightHand,LeftFoot;RighFoot;Head	61,80

observations for rotational data [1]. The analysis of displacements of the LeftUpLeg and RightUpLeg markers allows to identify gaits with almost 85% accuracy, which means only nine more missclassified instances in comparing to complete set of markers.

There are still possibilities to improve the performance of the classification. Similar as in [1] the method can be extended by an automatic selection stage with usage of chosen heuristics. It would explore the pose space more accurately. What is more in equation (2) weight coefficients can be introduced [9, 10], which allow to specify an importance of successive markers. The problem of their determination is even more challenging in comparing to binary selection, however once again some heuristics can be applied. Another improvement may be related to an analysis of temporal context of virtual markers positions by extending cloud points with data of some preceding poses. [9, 10].

ACKNOWLEDGMENTS

This work has been supported by the National Centre for Research and Development (project UOD-DEM-1-183/001 "Intelligent video analysis system for behavior and event recognition in surveillance networks").

REFERENCES

1. A. Switonski, H. Josinski, H. Zghidi, and K. Wojciechowski, “Selection of pose configuration parameters of motion capture data based on Dynamic Time Warping,” in *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON NUMERICAL ANALYSIS AND APPLIED MATHEMATICS 2014*, American Institute of Physics, New York, 2015, vol. 1648 of *AIP Conference Proceedings*.
2. T. Krzeszowski, A. Switonski, B. Kwolek, H. Josinski, and K. Wojciechowski, *Lecture Notes in Computer Science* **8671**, 356–363 (2014).
3. H. Sakoe, and S. Chuba, *IEEE Transactions on Acoustics, Speech and Signal Processing* **8**, 43–49 (1978).
4. H. Andre-Jonsson, and D. Z. Badal, *Principles of Data Mining and Knowledge Discovery -Lecture Notes in Computer Science* **1263**, 211–220 (1997).
5. L. Chen, M. T. ÃŰzsu, and V. Oria, “Robust and fast similarity search for moving object trajectories,” in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, 2005, pp. 491–502.
6. M. D Morse, and J. M. Patel, “An efficient and accurate method for evaluating time series similarity,” in *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 2007, pp. 569–580.
7. L. Chen, and N. Raymond, “On the marriage of Lp-norms and edit distance,” in *VLDB ’04 Proceedings of the Thirtieth international conference on Very large data bases*, 2004, vol. 30, pp. 792–803.
8. X. Wang, H. Ding, G. Trajcevski, P. Scheuermann, and E. J. Keogh, *Data Mining and Knowledge Discovery* **26**, 275–309 (2013).
9. L. Kovar, M. Gleicher, and P. Fredeeic, *ACM Transactions on Graphics* **21**, 473–482 (2002).
10. L. Kovar, and M. Gleicher, “Flexible automatic motion blending with registration curves,” in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003, pp. 214–224.
11. T. Roder, *Similarity, Retrieval, and Classificationof Motion Capture Data*, Ph.D. thesis, Rheinischen Friedrich-Wilhelms-Universitat (2007).

12. I. Witten, E. Frank, and M. Hall, *Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Boston, Massachusetts, 2011, ISBN 978-0-12-374856-0.