I.J. Intelligent Systems and Applications, **2016, 6, 1-8**
Published Online June 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2016.06.01

An Approach to Gesture Recognition with Skeletal Data Using Dynamic Time Warping and Nearest Neighbour Classifier

Alba Ribó
University of Lleida, 25001, Catalonia, Spain
E-mail: albaribopalenzuela@gmail.com

Dawid Warchoł
Department of Computer and Control Engineering, Rzeszow University of Technology
W. Pola 2, 35-959 Rzeszow, Poland
E-mail: dawwar@kia.prz.edu.pl

Mariusz Oszust
Department of Computer and Control Engineering, Rzeszow University of Technology
W. Pola 2, 35-959 Rzeszow, Poland
E-mail: marosz@kia.prz.edu.pl

***Abstract*—**Gestures are natural means of communication between humans, and therefore their application would benefit to many fields where usage of typical input devices, such as keyboards or joysticks is cumbersome or unpractical (e.g., in noisy environment). Recently, together with emergence of new cameras that allow obtaining not only colour images of observed scene, but also offer the software developer rich information on the number of seen humans and, what is most interesting, 3D positions of their body parts, practical applications using body gestures have become more popular. Such information is presented in a form of skeletal data. In this paper, an approach to gesture recognition based on skeletal data using nearest neighbour classifier with dynamic time warping is presented. Since similar approaches are widely used in the literature, a few practical improvements that led to better recognition results are proposed. The approach is extensively evaluated on three publicly available gesture datasets and compared with state-of-the-art classifiers. For some gesture datasets, the proposed approach outperformed its competitors in terms of recognition rate and time of recognition.

***Index Terms*—**Gesture Recognition, Nearest Neighbour Classifier, Dynamic Time Warping, Kinect, Skeletal Data, Matlab.

I. Introduction

The subject of gesture recognition has been widely explored through the last two decades. The main reason of such interest is omnipresence of gestures in daily life. They are mostly used to support verbal communication, but, in many cases where spoken language cannot be used, they are indispensable. As a natural mean of communication, gestures are very demanded in human-machine interaction field. Their present applications involve video games industry or home automation. Gesture recognition technology may help people with hearing disabilities to be understood. Sign language recognition, as a challenging example of gesture recognition problem, is also the subject of many recent approaches, e.g., [1-7]. Here, approaches handle large gesture variation, i.e., gesture's executions from a given class can be performed with different speeds, and there is a dependency of gesture execution on signer's anatomical constraints or intention. Furthermore, each nationality has its own sign language.

Another direction of research is associated with recognition of gestures performed by entire body [8, 9], with some attention to assistive technology [9-11].

There are also works using additional equipment, such as gloves [12] or surface electrodes [13]. They provide accurate information on hand positions but may interfere in the proper gesture execution.

Other common approaches have turned towards computer vision, in which gesture recognition is based on colour information [14]. Neural networks [4, 15], hidden Markov models [2, 7, 16, 17], support vector machines [16], boosting [14] or nearest neighbour with dynamic time warping (DTW) [1, 2, 16, 18] are mostly used as classifiers.

Since DTW with nearest neighbour seems to be one of the simplest approaches, which is also able to obtain state-of-the-art results, we extended this approach adding a few practical improvements and evaluated the resulting classifier on three widely used gesture datasets. The results were compared with results obtained by state-of-the-art approaches.

The rest of the paper is organised as follows. Section II

describes related works on gesture recognition. Section III gives background information concerning Kinect sensor, DTW and nearest neighbour classifier. In Section IV, our approach is presented, evaluated and compared with representative approaches. Section V concludes the paper.

II. Related Works

The recently developed methods of gesture recognition include the use of active depth cameras or sensors, which provide substantial data from the observed environment, such as three-dimensional information. Kinect sensor [19] is an example of such device. It provides colour image, depth map, and 3D skeletal data indicating the most important 20 body joints. Among works with Kinect, Lai *et al.* [20] recognised eight static hand gestures with accuracy of 99%, and Ren *et al.* [21] recognised 14 static hand shapes which were controlling an application performing arithmetic operations, and also three shapes for *Rock-paper-scissors* game. Recognition experiments with simple body or hand movements can be found in works [22, 23]. In [1-7, 24-25], in turn, sign language gestures were recognised using skeletal data. In [26], only depth maps obtained from the sensor were used.

Since sign language recognition is very challenging, it has attracted many researchers, and therefore some of recently developed, significant solutions that are using Kinect are worth to be presented. For example, in [1] time series characterising isolated Polish Sign Language words were classified. In [2], in turn, authors used depth data and Viewpoint Feature Histogram as the global descriptor of the scene for Polish and American static and dynamic hand gestures recognition. An American Sign Language was recognised in work of Sun et al. [3], where a latent support vector machine model was developed for sign classification. Authors utilised colour information, as well as depth and skeletal data. Convolutional neural network was applied for feature construction process in [4]. In that work, 20 Italian gestures were recognised with 91.7% accuracy. What is important, described approach achieved a mean Jaccard index of 0.789 in the ChaLearn 2014 Looking at People gesture spotting competition. Halim and Abbas in [5] presented DTW-based approach for Pakistani Sign Language recognition with accuracy of 91%. Zafrulla et al. in [6] presented American Sign Language recognition dataset and an approach based on random forest regression. The approach used depth images in order to improve Microsoft Kinect Skeleton Tracker. Yang in [7] presented hierarchical conditional random fields applied for detection of signs' segments and verified hand shapes of the segmented signs using BoostMap. Hand shapes have been also used in work [8], where a new superpixel earth mover's distance metric, depth and skeletal data together outperformed compared approaches in real-life examples. Jiang et al. in [9] proposed a multi-layered gesture recognition method with Kinect, which was able to obtain promising results in the one-shot learning gesture recognition test on ChaLearn gesture dataset.

Apart from sign language recognition, there are also interesting approaches covering other aspects of human life. For example, authors in [10] presented a system that recognises gestures using Kinect for elderly to call service robot for their service request. The approach also utilises face detection and skin colour information, which are particularly useful when skeletal data is not available.

Another approach is applied to automatic diet monitoring system for elderly [11]. In [27], authors proposed using Kinect for rehabilitation of patients with Metachromatic leukodystrophy. For further reading, a thorough review of the recent works with Kinect and their impact on physical therapy and rehabilitation can be found in [28].

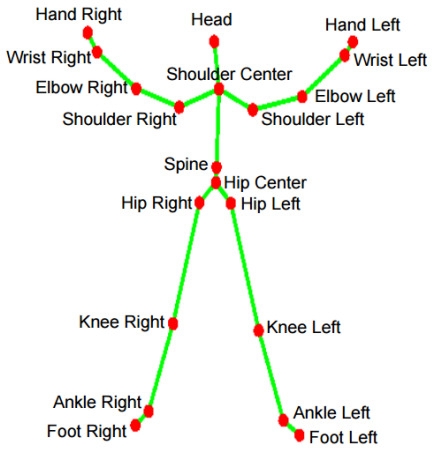

Fig.1. Kinect Joints [30]

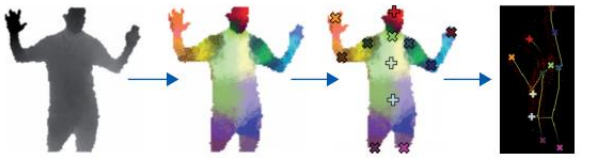

Fig.2. Skeletal Tracking Pipeline in Kinect [19]

III. Background

This section begins with introduction of a sensor, which was used in order to obtain experimental data. Since the proposed approach covers a subject of gesture recognition, some preliminary information on used distance metric and a method of classification with this metric is also given.

A. Kinect

As described in [19, 29], Kinect contains a depth sensor (infrared projector and a camera) and a colour camera. It analyses the deformation of a known speckle pattern of infrared laser light when it is projected onto the scene and finally constructs a depth map. The one of the most appreciated Kinect's advantages is its ability to track skeletal movements in real time. Figure 1 presents

20 joints which represent the human body in the skeletal tracking. In order to obtain the skeleton, at first body parts are recognised, and then the body joints are hypothesized and mapped to a skeleton using temporal continuity and prior knowledge (see Figure 2). The newly developed version of Kinect also offers a neck joint and a differentiation of hand tip and thumb joints.

B. Dynamic Time Warping

Dynamic time warping (DTW) is a dynamic programming technique for measuring similarities between two time series that may differ in speed or acceleration. It has its origins in speech recognition.

In the algorithm, two time series of lengths s and t are processed. Here, the s-by-t matrix is generated by calculation of distances between their consecutive samples. Distances are calculated with either Euclidean metric or city-block metric. Then, a so-called warping path is created. The path must satisfy three following conditions: boundary, continuity and monotonicity. The boundary condition constraint requires the warping path to start and finish in diagonally opposite corner cells of the matrix. The continuity constraint restricts the allowable steps to adjacent cells. The monotonicity constraint forces the points in the warping path to be monotonically arranged in time. The summed values in cells along the shortest path are returned as the DTW distance between compared time series.

To speed up the process, a window constraint (window size, W) is introduced. In this case, instead of comparing each n-th sample of the first time series to all the samples of the second time series, the first one is compared with the samples from $n-W$ to $n+W$ of the second time series. If the specified window size is smaller than the difference between lengths of the time series ($W < |s-t|$), then W is set equal to this difference

C. K-Nearest Neighbour Classifier

K-nearest neighbour classifier (KNN) is used to assign tested objects comparing them with labelled classes' representatives. The tested object obtains the label of the closest K representatives from a given class. In our case, DTW is the distance used to compare time series composed of frames of skeletal data.

IV. The Approach and Experimental Evaluation

Since DTW and KNN have already been applied to gesture recognition with Kinect, we introduce several practical improvements starting from finding appropriate W in DTW, as well as experimenting with reduced number of processed joints. The proposals have a practical nature and, as it is shown in the experimental part of this section, they allow obtaining better results than some of the state-of-the-art techniques. In experiments, we reduced the number of joints leaving only arm joints (right and left elbow, wrist and hand), as they are thought to be the only ones that contribute. The experiments presented in this section were executed using Matlab R2014a on a machine with i7-4510U CPU and 8GB RAM.

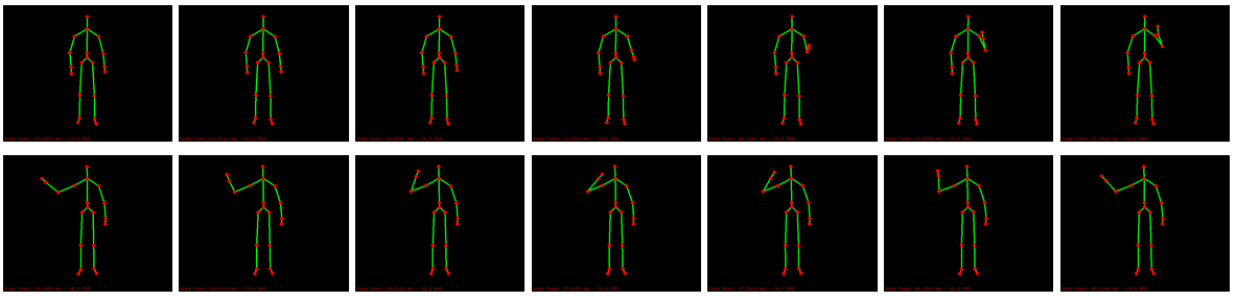

Fig.3. Skeletal images that belongs to two exemplary gestures from *VISAPP2013* dataset [30] (one gesture in a row)

Table 1. 10-fold cross-validation tests on *Visapp2013*; DTW parameters: metric and window size (W); KNN parameter: K; *Inf* denotes infinity; results in [%]

	Arm Joints										**All Joints**	
DTW	Euclidean metric					City-block metric					City-block metric	
K	W=1	W=3	W=5	W=10	W=*Inf*	W=1	W=3	W=5	W=10	W=*Inf*	W=10	W=*Inf*
1	95.45	95.45	95.45	95.45	95.45	95.91	95.91	95.91	95.91	95.91	92.73	92.73
3	94.55	94.55	94.55	94.55	94.55	95.45	95.45	95.45	95.45	95.45	93.64	93.64
5	94.55	94.55	94.55	94.55	94.55	96.36	96.36	96.36	96.36	96.36	93.64	93.64
7	96.36	96.36	96.36	96.36	96.36	97.27	97.27	97.27	97.73	97.73	98.18	98.18
9	99.09	99.09	99.09	99.55	99.55	99.09	99.09	99.55	100	100	96.82	96.82

Table 2. Confusion matrix for gesture classification for Visapp2013 dataset, with the best DTW and KNN parameters

	1	2	3	4	5	6	7	8	Recognition rate, in [%]
LH Pull D - 1	17				3				85
LH Push U - 2		17				3			85
LH Swipe R - 3			14		6				70
LH Wave - 4	2			18					90
RH Pull D - 5					20				100
RH Push U - 6						20			100
RH Swipe Lt - 7	2						18		90
RH Wave - 8					1			19	95

In experiments, we have performed 10-fold cross-validation tests with KNN and DTW. In these tests, the benchmark dataset is divided into ten subsets, and then nine of them are used as a training test, while the remaining subset is used as a test set.

We used three datasets for evaluation; the first one comes along with predefined training and test sets. *Visapp2013* dataset [30] contains eight gestures, each performed eight times in the original training set and 20 times in the original test set. On the whole, eight gestures were performed 28 times, resulting in the total number of gestures equal to 224. Two exemplary gestures are presented on Figure 3. In 10-fold cross-validation test, the last four realisations were not used.

For the first set of tests, fifty 10-fold cross-validation tests were performed changing metrics in DTW between Euclidean and city-block, as well as the window size W = 1, 3, 5, 10 or infinite. KNN requires determination of K parameter; therefore it was set to 1, 3, 5, 7, and 9.

The approach was tested using the remaining datasets and the best set of parameters (city-block metric, W = 10 or infinite, and K = 9). The usage of arm joints yielded faster and more accurate recognition. The results for *Visapp2013* can be found in Table 1.

The test considering arm joints not only lead to better results but also turned out to be much less time consuming (548s) than the one involving all joints (1815s).

For the following tests, the DTW was used with the city-block metric and W = 10. The results of the test matching the original test set with the original training set offer best results for K = 9, reaching 89.38% accuracy. The confusion matrix associated with this dataset can be found in Table 2. From there we obtain gestures involving right arm that were better classified than those performed with the left arm. 100% of accuracy is reached for gestures 5 and 6 (*right hand pull down* and *right hand push up*), but three out of the 20 testing executions for those same gestures performed with the left hand were wrongly classified as the right hand. *Also left hand swipe right* has found to be similar to *right hand pull down* in six cases.

In Table 3 the comparison of the results with different K in the KNN classifications is given, as well as the results for other methods from the literature [30]. It is worth noticing that in these works only six out of the eight gestures were considered.

Table 3. Recognition rate comparison for *Visapp2013* dataset; results in [%]

Method	Recognition rate
Classical DTW [30]	60.0
Weighted DTW 1 [30]	62.5
Weighted DTW 2 [30]	96.7
DTW & 1-NN Classifier	88.7
DTW & 3-NN Classifier	86.9
DTW & 5-NN Classifier	86.2
DTW & 7-NN Classifier	88.7
DTW & 9-NN Classifier	89.4

The approach presented in this paper outperformed the results obtained with classical DTW by 29%, and the weighted DTW by 27%, all with K = 9. The second approach with DTW is ca. 7% more accurate.

MSR Action Screen Coordinates and *MSR Action Real World Coordinates* are the remaining gesture databases. They contain the same realisations of gestures but they are stored with screen and real world coordinates, respectively. The gestures involve all body movements, not just arms. From this reason, we take into consideration all recorded joints. For these two databases, the subsamples for the 10-fold cross-validation were organised in the same way and equivalently to the first database. They contain 20 actions performed by ten subjects and they were executed three times. There are 600 realisations in total. However, in some actions, one or more subjects are missing and also in some few cases there are only two executions, what reduces the total number of realisations to 567. In cross-validation, the last seven executions were skipped, and thus, each subset contains 56 realisations.

The 10-fold cross-validation results for MSR Action Screen Coordinates and Real World Coordinates are given in Table 4.

Table 4. 10-fold cross-validation tests on MSR Action dataset; results in [%]

DTW	City-block metric, window size (W) = 10	
K	Screen Coordinates	Real World Coordinates
1	94.11	95.71
3	88.93	91.43
5	85.71	90.00
7	87.14	90.00
9	85.71	88.57

In all the cases, the usage of the real world coordinates led to better results. Opposite to the first dataset, the best result of *MSR Action* database is obtained with 1-nearest neighbour classifier (95.71% correctly recognised gestures) and the precision of the classification results tend to decrease while increasing the *K* in KNN. The confusion matrix for the best performance is presented in Table 5 for *Screen Coordinates* and Table 6 for *Real World Coordinates*.

Table 5. Confusion matrix for 10-fold cross-validation tests on *MSR Action Screen Coordinates* dataset; *K* = 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Recognition rate, in [%]
High Arm Wave - 1	23	3										1									27	85
Horizontal Arm Wave - 2		25	1	1																	27	93
Hammer - 3			27																		27	100
Hand Catch - 4			1	20	3		1					1									26	77
Forward Punch - 5					25	1															26	96
High Throw - 6				2		21					1							2			26	81
Draw X - 7							25	1	1				1								28	89
Draw Tick - 8								30													30	100
Draw Circle - 9	1								29												30	97
Clap Front - 10										30											30	100
Two Hand Wave - 11						2				1	27										30	90
Side Boxing - 12												30									30	100
Bend - 13							1					1	26							2	30	87
Front Kick - 14													1	28							29	97
Side Kick - 15															19						19	100
Jogging - 16																29					29	100
Tennis Swing - 17												1					27	1			29	93
Tennis Serve - 18						1						1						27			29	93
Golf Swing - 19																			29		29	100
Pick up And Throw - 20													1							28	29	97

Table 6. Confusion matrix for 10-fold cross-validation on MSR Action Real World Coordinates dataset; K = 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Recognition rate, in [%]
High Arm Wave - 1	23	3										1									27	85
Horizontal Arm Wave - 2		25	1		1																27	93
Hammer - 3			27																		27	100
Hand Catch - 4			1	22	1		1					1									26	85
Forward Punch - 5					26																26	100
High Throw - 6				1		25															26	96
Draw X - 7							25	1	1				1								28	89
Draw Tick - 8								30													30	100
Draw Circle - 9	1								29												30	97
Clap Front - 10										30											30	100
Two Hand Wave - 11										1	29										30	97
Side Boxing - 12												30									30	100
Bend - 13							1					1	27							1	30	90
Front Kick - 14													1	28							29	97
Side Kick - 15															19						19	100
Jogging - 16																29					29	100
Tennis Swing - 17												1					27	1			29	93
Tennis Serve - 18												1						28			29	97
Golf Swing - 19																			29		29	100
Pick up And Throw - 20													1							28	29	97

It is worth noting that those two datasets contain information of the same gestures executions stored in a different way, as the confusion matrices obtained hold a great similarity with slight improvements in the second one. Both present a high recognition rate, reaching 100% in many cases. *High arm wave* is found similar to *horizontal arm wave* in three out of the 27 executions. *Hand catch* is confused with other four actions: three times in particular with *forward punch* in screen coordinates and once in real world coordinates.

Other works [31, 32] have been using raw depth maps from these databases. In [31], an action graph was used to

model each action. It was constructed using the concept of bag of points. They perform experiments splitting the whole dataset into three different sets containing eight gestures each (some were repeated) and three tests were performed: (1) using 1/3 of the samples as training set, (2) having 2/3 of the samples as training, and (3) doing a cross-subject test using 1, 3, 5, 7, 9 subjects as training and 2, 4, 6, 8, 10 as testing. Results are given in Table 7.

In [32], a novel features and actionlet ensemble model for human gestures recognitions from depth map were proposed. The results for the average of the 252 possible 5-5 cross-subject tests are compared with other state-of-the-art methods in Table 8. Cross-subject tests were also performed with our proposed approach using the following subjects for training: 1, 3, 5, 7, 9, and subjects: 2, 4, 6, 8, and 10 for testing. The approach correctly recognised 68% of gestures with K = 1, and 68.4% with K = 9. The proposed method using city-block metric, W = 10 and K = 9 in KNN yielded better results than previously obtained with DTW by 14.4%. The obtained results are not better than state-of-the-art approaches, e.g., the ones using actionlets [32], however, some of the proposed improvements led to better results with DTW than the results with DTW reported in the literature.

Table 7. Results on *MSR Action* dataset [31] (in [%])

	Test 1	Test 2	Test 3
Set 1	89.5	93.4	72.9
Set 2	89.0	92.9	71.9
Set 3	96.3	96.3	79.2
Overall	91.6	94.2	74.7

Table 8. Cross-subject tests with state-of-the-art methods on *MSR Action* dataset; results in [%]

Method	Recognition rate
Recurrent Neural Network [33]	42.5
Dynamic Temporal Warping [34]	54.0
Hidden Markov Model [35]	63.0
Action Graph on Bag of 3D Points [31]	74.7
Actionlet Ensemble [32]	88.2
Proposed Approach, K=1	68.0
Proposed Approach, K=9	68.4

V. Conclusion

Recent development of new sensors that allow tracking important parts of the human body resulted in proliferation of different approaches to gesture recognition and their practical applications. Therefore, any improvement of existing solutions towards better recognition accuracy or shortening recognition time is very important. In this paper, we proposed several improvements: reducing processed number of joints, investigating the K in KNN, or W in DTW. The proposed approach turned out to be better than some of the compared state-of-the-art techniques.

In future works, we plan to reduce training sets choosing appropriate classes' representatives in order to reduce the recognition time. Here, proximity matrices that contain DTW distances could be in use [1]. Another research direction could involve comparing accuracy of the presented approach with results of other classifiers, such as neural networks or support vector machines [36].

Acknowledgment

We would like to show our gratitude to IAESTE (The International Association for the Exchange of Students for Technical Experience) for providing the opportunity of working together with an enthusiastic and capable team. Also the members of the IAESTE Local Committees from Rzeszow University of Technology and Polytechnic School of University of Lleida, as well as the management of the Department of Computer and Control Engineering of Rzeszow University of Technology, for making it possible.

Author Contributions

MO and DW designed experiments. AR implemented the approach and performed experiments. AR, MO, and DW conceived of the study, performed the data analysis and drafted the manuscript. All authors read and approved the final version of the manuscript.

References

[1] M. Oszust, and M. Wysocki, "Recognition of Signed Expressions Observed by Kinect Sensor," *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference*, pp. 220-225, 2013, doi: 10.1109/AVSS.2013.6636643.

[2] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchoł, "Recognition of Hand Gestures Observed by Depth Cameras," *International Journal of Advanced Robotic Systems*, January 2015, doi: 10.5772/60091.

[3] C. Sun, T. Zhang, and C. Xu, "Latent Support Vector Machine Modeling for Sign Language Recognition with Kinect," *ACM Trans. Intell. Syst. Technol.*, 6(2), 2015, doi: 10.1145/2629481.

[4] L. Pigou, S. Dieleman, P. J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks", *European Conference on Computer Vision (ECCV) Workshops, Lecture Notes in Computer Science*, vol. 8925, pp. 572-578, Springer, 2015, doi: 10.1007/978-3-319-16178-5_40.

[5] Z. Halim and G. Abbas, "A Kinect-Based Sign Language Hand Gesture Recognition System for Hearing- and Speech-Impaired: A Pilot Study of Pakistani Sign Language," *Assistive Technology: The Official Journal of RESNA*, pp. 34-43, 2014, doi: 10.1080/10400435.2014.952845.

[6] Z. Zafrulla, H. Sahni, A. Bedri, P. Thukral, and T. Starner, "Hand detection in American Sign Language depth data using domain-driven random forest regression," *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp. 1-7, 2015, doi: 10.1109/FG.2015.7163135.

[7] H. D. Yang, "Sign language recognition with the Kinect sensor based on conditional random fields," *Sensors*, 15(1), 135-147, 2015, doi: 10.3390/s150100135.

[8] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-Based Hand Gesture Recognition with Kinect Depth Camera," *IEEE Transactions on Multimedia*, 17(1), pp. 29-39, 2015, doi: 10.1109/TMM.2014.2374357.

[9] F. Jiang, S. Zhang, S. Wu, Y. Gao, and D. Zhao, "Multi-layered Gesture Recognition with Kinect," *Journal of Machine Learning Research*, vol. 16, pp. 227-254, 2015.

[10] Z. Xinshuang, A. M. Naguib, and S. Lee, "Kinect based calling gesture recognition for taking order service of elderly care robot," *23rd IEEE International Symposium on Robot and Human Interactive Communication RO-MAN*, pp. 525-530 2014, doi: 10.1109/ROMAN.2014.6926306.

[11] A. Cunhaa, L. Páduab, L. Costab, and P. Trigueiros, "Evaluation of MS Kinect for Elderly Meal Intake Monitoring," *6th Conference on ENTERprise Information Systems– aligning technology, organizations and people, CENTERIS*, vol. 16, pp. 1383–1390, 2014, doi: 10.1016/j.protcy.2014.10.156.

[12] C. Vogler and D. N. Metaxas, "Toward Scalability in ASL Recognition: Breaking Down Signs into Phonemes," A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil editors, *Gesture Workshop*, vol. 1739, *Lecture Notes in Computer Science*, pp. 211–224. 1999, doi: 10.1007/3-540-46616-9_19.

[13] Gupta, S. Kundu, R. Pandey, R. Ghosh, R. Bag, and A. Mallik, "Hand Gesture Recognition and Classification by Discriminant and Principal Component Analysis Using Machine Learning Techniques," *IJARAI*, 1(9), 2012.

[14] G. Awad, J. Han, and A. Sutherland, "Novel Boosting Framework for Subunit-Based Sign Language Recognition," *Proceedings of the 16th IEEE international conference on Image processing*, pp. 2693–2696, 2009, doi: 10.1109/ICIP.2009.5414159.

[15] M. Zahedi and A. R. Manashty, "Robust Sign Language Recognition System Using Tof Depth Cameras," *Computing Research Repository - CORR*, abs/1105.0, 2011.

[16] V. Dixit, and A. Agrawal, "Real Time Hand Detection & Tracking for Dynamic Gesture Recognition," *IJISA*, 7(8), pp. 38-44, 2015, doi: 10.5815/ijisa.2015.08.05.

[17] W. Gao, G. Fang, D. Zhao, and Y. Chen. "A Chinese Sign Language Recognition System Based on SOFM/SRN/HMM," *Pattern Recognition*, 37(12), pp. 2389–2402, 2004, doi: 10.1016/j.patcog.2004.04.008.

[18] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, R. Stefan, A. Thangali, H. Wang, and Q. Yuan. "Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/ Retrieval Algorithms," *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT)*, pp. 11-14, 2010.

[19] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, *et al.* "Real-time Human Pose Recognition in Parts From Single Depth Images," *Communications of the ACM*. 56(1), pp. 116-124, 2013, doi: 10.1007/978-3-642-28661-2_5.

[20] K. Lai, J. Konrad, and P. Ishwar, "A Gesture-Driven Computer Interface Using Kinect," *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium*, pp. 185–188, 2012, doi: 10.1109/SSIAI.2012.6202484.

[21] Z. Ren, J. Meng, J. Yuan, and Z. Zhang. "Robust Hand Gesture Recognition with Kinect Sensor," *Proceedings of the19th ACM international conference on Multimedia*, pp. 759–760, ACM, 2011, doi: 10.1145/2072298.2072443.

[22] K. K. Biswas and S. Basu,"Gesture Recognition Using Microsoft Kinect," *Automation, Robotics and Applications(ICARA), 2011 5th International Conference*, pp. 100–103, 2011, doi: 10.1109/ICARA.2011.6144864.

[23] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human Gesture Recognition Using Kinect camera," *Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference*, pp. 28–32, 2012, doi: 10.1109/JCSSE.2012.6261920.

[24] S. Lang, M. Block, and R. Rojas, "Sign Language Recognition Using Kinect," L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada editors, *Artificial Intelligence and Soft Computing*, vol. 7267, *Lecture Notes in Computer Science*, Springer, pp. 394–402, 2012, doi: 10.1007/978-3-642-29347-4_46.

[25] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti. "American Sign Language Recognition with the Kinect," *Proceedings of the 13th international conference on multimodal interfaces*, pp. 279–286, ACM, 2011, doi: 10.1145/2070481.2070532.

[26] D. Uebersax, J. Gall, M. Van den Bergh, and L. Van Gool, "Real-time sign language letter and word recognition from depth data," *Computer Vision Workshops (ICCV) Workshops, 2011 IEEE International Conference*, pp. 383–390, 2011, doi: 10.1109/ICCVW.2011.6130267.

[27] A. M. Ulaşlı, U. Türkmen, H. Toktas, and O. Solak, "The Complementary Role of the Kinect Virtual Reality Game Training in a Patient With Metachromatic Leukodystrophy," *PM&R*, 6(6), pp. 564-7, 2014, doi: 10.1016/j.pmrj.2013.11.010

[28] H. M. Hondori and M. Khademi, "A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation," Journal of Medical Engineering, vol. 2014, Article ID 846514, 16 pages, 2014. doi:10.1155/2014/846514.

[29] Z. Zhang, "Microsoft Kinect Sensor and its Effect." *Multimedia. IEEE*, 19(2), pp. 4-10, 2012, doi: 10.1109/MMUL.2012.24.

[30] S. Celebi, A. S. Aydin, T. T. Temiz, and T. Arici, "Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping," *VISAPP 2013* (1), pp. 620-625. 2013, doi: 10.5220/0004217606200625.

[31] W. Li, Z. Zhang, and Z. Liu "Action Recognition Based on A Bag of 3D Points," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference*. pp. 9-14, 2010, doi: 10.1109/CVPRW.2010.5543273.

[32] J. Wang, Z. Liu, Y. Wu, AND J. Yuan, "Mining Actionlet Ensemble for Action Recognition with Depth Cameras," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference*. pp. 1290-1297, 2012, doi: 10.1109/CVPR.2012.6247813.

[33] J. Martens and I. Sutskever, "Learning Recurrent Neural Networks with Hessian-Free Optimization," *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 2011.

[34] M. Muller, and T. Roder, "Motion Templates for Automatic Classification and Retrieval of Motion Capture Data," *Proceedings of the 2006 ACM SIGGRAPH/ Eurographics symposium on Computer animation*, pp. 137–146, Eurographics Association, 2006.

[35] F. Lv and R. Nevatia. "Recognition and Segmentation of 3-D Human Action Using HMM and Multi-class AdaBoost," *European Conference on Computer Vision (ECCV)*, pp. 359–372, 2006, doi: 10.1007/11744085_28.

[36] Saloni, R. K. Sharma, and Anil K. Gupta, "Voice Analysis for Telediagnosis of Parkinson Disease Using Artificial Neural Networks and Support Vector Machines," *IJISA*, 7(6), pp. 41-47, 2015, doi: 10.5815/ijisa.2015.06.04.

Authors' Profiles

Alba Ribó started the Degree in Industrial Electronics and Automatic Engineering of the University of Lleida in 2012. At summer 2015 she worked as an intern in the Department of Computer and Control Engineering of Rzeszow University of Technology, Poland. Her research interests include human-machine interaction and robotics.

Dawid Warchoł received his MSc in Computer Science in 2013 from the Rzeszow University of Technology, Poland. He is currently a doctoral candidate and research assistant in Department of Computer and Control Engineering, Rzeszow University of Technology, Poland. His research interests include multimedia, image processing, computer vision and pattern recognition.

Mariusz Oszust received the MSc degree in electrical engineering from the Rzeszow University of Technology in 2005 and PhD in computer science in 2013 from AGH University of Science and Technology, Krakow, Poland. Currently he is an assistant professor at Rzeszow University of Technology. His main research interests include pattern recognition, optimisation and development of vision-based human-computer interfaces. He is a member of ACM.

How to cite this paper: Alba Ribó, Dawid Warchoł, Mariusz Oszust, "An Approach to Gesture Recognition with Skeletal Data Using Dynamic Time Warping and Nearest Neighbour Classifier", International Journal of Intelligent Systems and Applications (IJISA), Vol.8, No.6, pp.1-8, 2016. DOI: 10.5815/ijisa.2016.06.01

