Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 571–575

Dependence of Kinect sensors number and position on gestures recognition with Gesture Description Language semantic classifier

Tomasz Hachaj
Pedagogical University of Krakow
2 Podchorazych Ave,
30-084 Krakow, Poland
Email: tomekhachaj@o2.pl

Marek R. Ogiela
AGH University of Science and Technology,
30 Mickiewicza Ave,
30-059 Krakow, Poland
Email: mogiela@agh.edu.pl

Marcin Piekarczyk
Pedagogical University of Krakow,
2 Podchorazych Ave,
30-084 Krakow, Poland
Email: marp@up.krakow.pl

***Abstract*—We have checked if it is possible to increase effectiveness of standard tracking library (Kinect Software Development Kit) by fusion of body joints gathered from different sensors positioned around the user. The proposed calibration procedure enables integration of skeleton data from set of tracking devices into one skeleton. That procedure eliminates many segmentation and tracking errors. The test set for our methodology was 700 recordings of seven various Okinawa Shorin-ryu Karate techniques performed by black belt instructor. In case when side Kinects were rotated in $\frac{\pi}{2}$ and -$\frac{\pi}{2}$ around vertical axis relatively to central one number of all not classified Karate techniques dropped by 48% while excessive misclassification error remained in the same level.**

***Index Terms*—Gesture recognition, Gesture Description Language, time sequence analysis, Kinect, pattern classification, semantic approach, Karate.**

I. Introduction

THE COMMON approach in gesture recognition is partitioning the movement sequence into sections that are represented by key frames. Those frames are than classified by different recognition techniques. For example in [1] authors propose an automatic learning method for gesture recognition. First, they apply the Self-Organizing Map to divide the sample data into phases and construct a state machine. Next, they apply the Support Vector Machine to learn the transition conditions between nodes. Nowadays multimedia devices that enable real-time tracking of observed users (like Microsoft Kinect controller) can be bought relatively cheaply. Because of that more and more researches apply them to record human body data (called body joints) that are automatically segmented from depth camera video data by dedicated software (like one implemented in Kinect SDK - Software Development Kit) like in [2], where a comparison of human gesture recognition using data mining classification methods in video streaming is proposed. The recognized gesture patterns of the study are stand, sit down, and lie down. Classification methods chosen for comparison study are back propagation neural network, support vector machine, decision tree, and naive Bayes. It has been proved that data acquired by Kinect device can be use not only to classify typical common - live gestures like waving or sitting but also to recognize high-speed gestures of martial arts sportsmen. In [3] authors aim at automatically recognizing sequences of complex Karate movements and giving a measure of the quality of the movements performed. The proposed system is constituted by four different modules: skeleton representation, pose classification, temporal alignment, and scoring. The proposed system is tested on a set of different punch, kick and defense Karate moves executed starting from the simplest case, i.e. fixed static stances up to sequences in which the starting stances is different from the ending one. All previously described methods uses many body features as an input for classification algorithm. However, in [4] the authors showed, that the majority of the information regarding the human motion resides in a lower dimensional space than one that can be obtained from all available features. These considerations further support the argument that human motion can be classified using a representation which considers a relatively low number of dimensions [5].

Knowing all of this we have proposed our new semantic classifier [6] called Gesture Description Language (GDL). The idea of GDL approach is to code the gesture sequences as the series of static key frames that appears in defined order. Those sequences are coded with context-free grammar called GDL script. GDL script consists of set of rules that creates together knowledge database similar to one an expert system has. The preliminary description of GDL architecture has been presented elsewhere [6], [7]. The basic assumption of GDL is:

- It is capable of classifying human body movements in real time.
- It can classify not only simple, real life gestures but also complicated movements like Karate techniques.
- It does not require large training dataset. Gestures are defined by user in GDL script. User can utilize as many body features as he or she needs in each rule definition.
- Gestures are split into key frames that appears in some order under given time restriction.
- The input data for classifier is set of body joints that arrive from tracking software in real - time (approximately with

978-1-4673-4471-5/$25.00

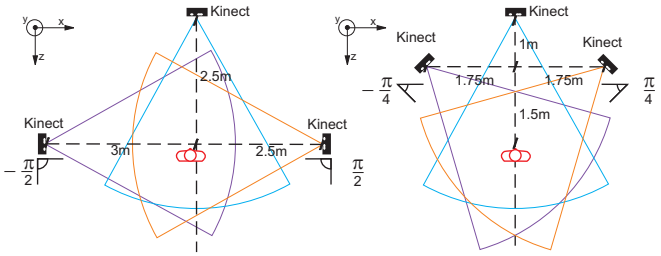

Fig. 1. Two tested multi-kinect environment configuration. Length of Kinect's view cone is 4 meters.

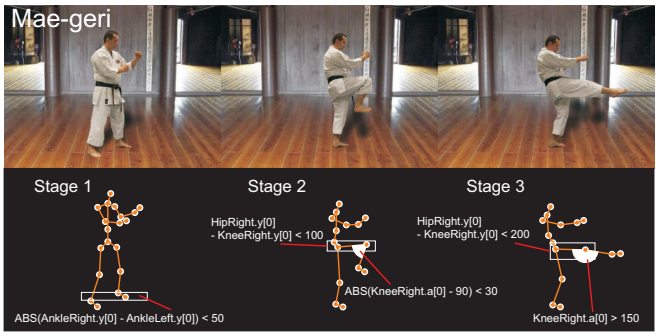

Fig. 2. Example Karate key-frames for GDL script. Movements are separated into stages, each stage is a key-frame used in semantic description. In this picture Mae-geri (front kick) begins with Moto-dachi (stance) but in our experimental recordings set it started from different, "neutral" position.

frequency 30 Hz). The set of tracked body joints is called Skelton (see Figure 2, bottom row).

- Our notation is invariant to user rotation around viewport of camera, because it can generates features regarding to angles measured between vectors defined by pairs of body joints (similarly to approach in [3]). Hovewer in opposite to [3] we can defines those angles dynamically while tailoring the GDL script description.

The idea of applying formal language to describe gestures is not new and was previously introduced for example in [8], [9]. However those papers describe only general framework of gesture description that might be potentially applicable for further recognition. The authors did not show how to use they annotations in pattern recognition tasks. They did not also validate they approach on any type of real-life data. Because of that those previous approaches were rather purely theoretical.

The novel contribution of this paper is test of GDL classifier performance on dataset that was acquired with one or three Kinect sensors that were positioned in two different configurations. We have checked if it is possible to increase effectiveness of standard tracking library (Kinect SDK) by fusion of body joints gathered from different sensors positioned around the user. The proposed calibration procedure enables integration of skeleton data from set of tracking devices into one skeleton. The test set was various Okinawa Shorin-ryu Karate techniques performed by black belt instructor. The whole solution runs in real-time and enables online and offline classification.

II. Material and Methods

In this section we will present experimental hardware setup, basis of GDL description and test dataset.

A. Multi - Kinect environment: setup and calibration

Figure 1 presents two tested multi-Kinect environment configuration. Each Kinect use its own tracking module (well know algorithm from Microsoft Kinect SDK which implementation can be used out of charges) that segments and tracks user skeleton in real time.

If front Kinect does not "see" particular body joint system checks if this joint is visible by another device. If yes our software takes coordinates measured by that second device. If more than two devices have detected same joint, coordinates are taken from camera that is closest to observed point. Each Kinect measure distance to observed point in its own right-handed Cartesian frame situated relatively to sensor orientation. Because of that same point V has different coordinates $\overline{v'} = [x', y', z', 1]$ and $\overline{v} = [x, y, z, 1]$ relatively to each pair of devices.

Our task now is to map all of those points to the same coordinate system. Let us assume that a Cartesian frame that represents orientation of each Kinect was translated and rotated around y (vertical) axis relatively to each other frame. That means there are four degrees of freedom (three for translation, one for rotation). Knowing that the linear transformation that maps coordinates of a point represented by vector $\overline{v'}$ in one coordinate system to coordinates $\overline{v}$ in another one has form of following matrix:

$$\overline{v'} \cdot \begin{bmatrix} cos(\beta) & 0 & -sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ sin(\beta) & 0 & cos(\beta) & 0 \\ t_x & t_y & t_z & 1 \end{bmatrix} = \overline{v} \tag{1}$$

In order to find unknown matrix coefficients following linear system has to be solved:

$$\begin{bmatrix} x'_1 & z'_1 & 1 & 0 \\ z'_1 & -x'_1 & 0 & 1 \\ x'_2 & z'_2 & 1 & 0 \\ z'_2 & -x'_2 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} cos(\beta) \\ sin(\beta) \\ t_x \\ t_z \end{bmatrix} = \begin{bmatrix} x_1 \\ z_1 \\ x_2 \\ z_2 \end{bmatrix} \tag{2}$$

$$y' + t_y = y \tag{3}$$

Where $\overline{v_1} = [x_1, y_1, z_1, 1]$, $\overline{v_2} = [x_2, y_2, z_2, 1]$ are points which coordinates are known in both frames. The linear system (2) and equation (3) is product of multiplication of matrix (1) by $\overline{v_1}$ and by $\overline{v_2}$. They multiplication by $\overline{v_1}$ produces the first and second equation in matrix (2) and equation (3). The multiplication by $\overline{v_2}$ produces third and fourth equation in matrix (2).

B. Bases of GDL scripts

The preliminary description of GDL architecture has been presented elsewhere [6], [7]. Because of that we will present only one example GDL script and its graphical explanation.

As we previously mentioned movement is separated into key frames. Each key frame is repressed by a rule that has a conclusion. If rule is satisfied for actual set of body joints positions (GDL uses forward chaining rezoning schema) its conclusion is memorized. It is possible to check with GDL script if some conclusion was satisfied in given time period. With this mechanism it is possible to generate key frames chains, which together create gesture. First row of Figure 2 presents three key frames of GDL script from Appendix that describes the Mae-geri kick. Second row explains body joints dependencies that are present in GDL script description. The last rule in script (the one with *sequenceexists* function) checks if all stages of movement have appeared in defined order under 1 second time restriction.

C. Test dataset

The dataset for our research are recordings of black belt Karate instructor[1] that performs seven different techniques: four static stances (Moto-dachi, Zenkutsu-dachi, Shiko-dachi and Naihanchi-dachi), two blocks (Gedan-uke and Age-uke) and one kick (Mae-geri). The instructor has indicated essential aspects of each technique (starting and ending positions of limbs and movement trajectory). The data was recorded during two sessions: one in which cameras was positioned as it was presented in Figure 1 on left, the second one as in Figure 1 on right. Second recording was done several weeks after the first one. Each gesture was partitioned into key-frames (Figure 3) that was later verified and accepted by instructor. Also expert was present during final validation of method. The same GDL script was used for all of recordings. The frame capture frequency was 30 Hz.

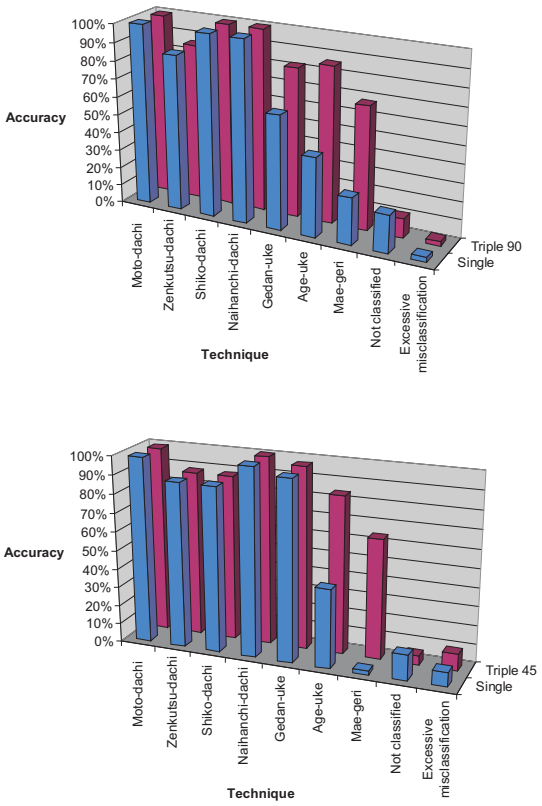

Fig. 3. Classification results from single and triple Kinect recordings. Triple 90 is left setup from Figure 1, Triple 45 is right setup from Figure 1.

III. Results

Tables 1-4 summarize the classification results of our experiment. The description in first column is the actual technique (or group of techniques) that is present in particular recording. Each technique (or group of techniques) was repeated 50 times. Symbol + means that particular recording consisted of more than one technique. Description in first row is classification results. Last but one row sums up percentage of correct classifications of particular technique. The last row sums up the percentage of correct classifications of techniques from first column. Summing up, we had 350 recordings of Karate techniques in each Kinect configuration (totally 700 recordings).

Because several Karate techniques can be present in same movement sequence we investigated if actual technique/techniques was/were classified. If yes that case was called correct classification. If technique was not classified and was not mistaken with similar one (like Moto-dachi which is similar to Zenkutsu-dachi) that case was called not classified. If technique was mistaken with similar one that case was called misclassified. Those three sums up to 100%. If technique was correctly classified but additional - actually not present - behavior was classified that case was called excessive misclassification. According to this terminology 90.4% of recordings from Table 4 was correctly classified, 5.2% was not classified and 4.4% was misclassified. Excessive misclassification was at the level of 9.0%. Figure 3 graphically presents results from Table 1-4.

IV. Discussion and conclusion

Our experiment has shown that integration of tracking data acquired by several Kinect devices with standard software increases the effectiveness of GDL classifier. This is due the fact that additional sensors that are situated at different angles than central one are capable of tracking body joints that in some situations might be covered by different body parts. This condition is especially visible in case of Mae-geri. Tracking of Karate kick is difficult task because of two factors: feet is moving with relatively high speed with large radius of path and in the last stage of Mae-geri feet is situated nearly at the same horizontal position as hip and knee. If the sportsman[2] is filmed only in front view knee and hip body joints are covered by feet and proper position of them have to be approximated by the software what, in practice,

[1]Karate instructor of Okinawa Shorin-ryu Karate with black belt degree (3 dan, sandan)

[2]In the meaning of Karate practitioner

TABLE I
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH SINGLE KINECT DEVICE (CENTRAL ONE) IN FIRST RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	6	0	0	0	7
Zenkutsu-dachi	1	37	0	0	1	0	0	12	1
Shiko-dachi +gedan-barai	0	0	50	0	27	0	0	0+23=23	0
Naihanchi-dachi	0	0	0	50	0	0	0	0	0
Gedan-barai +Zenkutsu-dachi	0	49	0	0	36	0	0	1+14=15	0
Age-uke +Moto-dachi	50	0	0	0	0	22	0	0+28=28	0
Mae-geri	4	0	11	0	0	0	13	26	4
%	100%	86.0%	100%	100%	63.0%	44.0%	26.0%	20.8%	2.4%

TABLE II
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH THREE KINECT DEVICES SITUATED AS SHOWN IN FIGURE 1 ON THE LEFT IN FIRST RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	6	0	0	0	7
Zenkutsu-dachi	1	37	0	0	1	0	0	12	1
Shiko-dachi +gedan-barai	0	0	50	0	46	0	0	0+4=4	0
Naihanchi-dachi	0	0	0	50	0	0	0	0	0
Gedan-barai +Zenkutsu-dachi	0	49	0	0	36	0	0	1+14=15	0
Age-uke +Moto-dachi	50	0	0	0	0	43	0	0+7=7	0
Mae-geri	4	0	0	0	0	0	34	16	4
%	100.0%	86.0%	100.0%	100.0%	82.0%	86.0%	68.0%	10.8%	2.4%

TABLE III
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH SINGLE KINECT DEVICE (CENTRAL ONE) IN SECOND RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	16	0	0	0	17
Zenkutsu-dachi	2	43	0	0	14	0	0	5	14
Shiko-dachi +gedan-barai	0	0	44	1	49	0	0	6+1=7	1
Naihanchi-dachi	0	0	0	50	7	0	0	0	7
Gedan-barai +Zenkutsu-dachi	2	45	0	0	47	0	0	3+3=6	0
Age-uke +Moto-dachi	49	0	0	0	0	21	0	1+29=30	0
Mae-geri	0	5	17	0	0	0	1	27	0
%	99.0%	88.0%	88.0%	100.0%	96.0%	42.0%	2.0%	15.0%	7.8%

TABLE IV
THE CLASSIFICATION RESULTS OF OUR EXPERIMENT. DATA WAS CAPTURED WITH THREE KINECT DEVICES SITUATED AS SHOWN IN FIGURE 1 ON THE RIGHT IN SECOND RECORDING SESSION.

	Moto-dachi	Zenkutsu -dachi	Shiko-dachi	Naihanchi -dachi	Gedan -barai	Age-uke	Mae-geri	Not classified	Excessive misclassification
Moto-dachi	50	1	0	0	16	0	0	0	17
Zenkutsu-dachi	2	43	0	0	14	0	0	5	14
Shiko-dachi +gedan-barai	0	0	44	1	49	0	0	6+1=7	1
Naihanchi-dachi	0	0	0	50	7	0	0	0	7
Gedan-barai +Zenkutsu-dachi	2	45	0	0	48	0	0	3+2=5	0
Age-uke +Moto-dachi	49	0	0	0	0	42	0	1+8=9	0
Mae-geri	0	0	23	1	0	0	32	0	6
%	99.0%	88.0%	88.0%	100.0%	97.0%	84.0%	64.0%	5.2%	9.0%

generates serious positioning errors. Our results have showed that in both configurations of multi-Kinect environment the effectiveness of classification increases because of increasing of tracking accuracy. In case when side Kinects were rotated about $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ around vertical axis relatively to central one number all not classified techniques dropped by 48% while excessive misclassification error remained on the same level. In case when Kinects were rotated about $\frac{\pi}{4}$ and $-\frac{\pi}{4}$ around vertical axis relatively to central one number all not classified techniques dropped by 61.9% while excessive misclassification error increased by 15.4%. It can be concluded that if we want to increase the correct classification factor in case when excessive misclassification error is not critical second setup of Kinects is more profitable. Otherwise, one should apply first setup, which, in our experiment did not change excessive misclassification rate.

Our future goal will be development of GDL script for recognition of complete set of most popular Karate techniques. The completed classifier will be than utilized in self-training multimedia application. We also plan to apply our classifier as a part of touchless interface in our medical data visualization module [10]. This will allow medical personnel to personally access patient data during surgical interventions while their hands are sterile. We also consider to expand GDL script terminal symbols and to test its capability in recognition of sign language gestures [11].

APPENDIX

The GDL script for Mae-geri recognition.

////////////////////
//Mae-geri
////////////////////
//Both legs are in the same level above the ground
//Figure 2 Mae-geri stage 1
RULE ABS(AnkleRight.y[0] - AnkleLeft.y[0]) $<$ 50
THEN MaeStart

//Right knee in the line with right hip, bended
//right knee
//Figure 2 Mae-geri stage 2
RULE (HipRight.y[0] - KneeRight.y[0]) $<$ 100
& ABS(KneeRight.a[0] - 90) $<$ 30
THEN MaeMiddleRight

//Kick with right foot - Figure 3 Mae-geri stage 3
RULE (HipRight.y[0] - KneeRight.y[0]) $<$ 200
& KneeRight.a[0] $>$ 150
THEN MaeEndRight

//Left knee in the line with left hip, bended left knee
//Figure 2 Mae-geri stage 2
RULE (HipLeft.y[0] - KneeLeft.y[0]) $<$ 100
& ABS(KneeLeft.a[0] - 90) $<$ 30
THEN MaeMiddleLeft

//Kick with left foot - Figure 3 Mae-geri stage 3
RULE (HipLeft.y[0] - KneeLeft.y[0]) $<$ 200
& KneeLeft.a[0] $>$ 150
THEN MaeEndLeft

//Proper sequence of Mae-geri stages
RULE (sequenceexists("[MaeMiddleRight,1][MaeStart,1]")
& MaeEndRight) |
(sequenceexists("[MaeMiddleLeft,1][MaeStart,1]")
& MaeEndLeft)
THEN Mae-geri

ACKNOWLEDGMENT

We kindly acknowledge the support of this study by a Pedagogical University of Krakow Statutory Research Grant.

REFERENCES

[1] M. Oshita, T. Matsunaga, Automatic Learning of Gesture Recognition Model Using SOM and SVM, Advances in Visual Computing, Lecture Notes in Computer Science, vol. 6453, 2010, pp. 751–759'
[2] O. Patsadu, C. Nukoolkit, B. Watanapa, Human gesture recognition using Kinect camera, Joint International Conference on Computer Science and Software Engineering (JCSSE), 2012, pp. 28–32.
[3] S. Bianco, F. Tisato, Karate moves recognition from skeletal motion, Proc. SPIE 8650, Three-Dimensional Image Processing (3DIP) and Applications 2013, 86500K (March 12, 2013); doi:10.1117/12.2006229.
[4] V. Ntouskos, P. Papadakis, F. Pirri, A Comprehensive Analysis of Human Motion Capture Data for Action Recognition, VISAPP 1, pp. 647–652. SciTePress, (2012).
[5] M. Trzupek, Semantic Interpretation of Heart Vessel Structures Based on Graph Grammars, Computer Vision and Graphics, Lecture Notes in Computer Science, vol. 6374, 2010, pp. 81–88.
[6] T. Hachaj, M. R. Ogiela, Recognition of human body poses and gesture sequences with gesture description language, Journal of medical informatics and technology, vol 20/2012, ISSN 1642-6037, October 2012, pp. 129–135.
[7] T. Hachaj, M. R. Ogiela, Semantic Description and Recognition of Human Body Poses and Movement Sequences with Gesture Description Language, Computer Applications for Bio-technology, Multimedia, and Ubiquitous City, Communications in Computer and Information Science, Vol. 353, 2012, pp. 1–8, Springer, Heidelberg.
[8] F. Echtler, G. Klinker, A. Butz, Towards a unified gesture description language, Proceeding HC '10 Proceedings of the 13th International Conference on Humans and Computers, pp. 177–182 University of Aizu Press Fukushima-ken, Japan, 2010.
[9] M. Kölsch, C. Martell, Towards a Common Human Gesture Description Language, Workshop on Mixed Reality User Interfaces, at VR 2006.
[10] T. Hachaj and M. R. Ogiela, Framework for cognitive analysis of dynamic perfusion computed tomography with visualization of large volumetric data, Journal of Electronic Imaging, vol. 21, issue 4, 10.1117/1.JEI.21.4.043017, 2012.
[11] W. Koziol, H. Wojtowicz, K. Sikora, W. Wajs, Analysis and Synthesis of the System for Processing of Sign Language Gestures and Translatation of Mimic Subcode in Communication with Deaf People, Knowledge Engineering, Machine Learning and Lattice Computing with Applications, Lecture Notes in Computer Science, vol. 7828, 2013, pp 61–70.