ICIC Express Letters
Part B: Applications
Volume **11**, Number **5**, May **2020**

 ISSN 2185-2766
pp. 487–492

A STUDY ON FINDING DIFFERENCES IN MOVEMENT OF EXPERT AND NOVICE DARTS PLAYERS BY USING A KINECT-LIKE 3D IMAGE SENSOR

Hajime Murao

Graduate School of Intercultural Studies
Kobe University
1-2-1, Tsurukabuto, Nada-ku, Kobe 657-8501, Japan
murao@i.cla.kobe-u.ac.jp

Received November 2019; accepted February 2020

Abstract. *What is an essential difference between the expert movement and the novice one where the movement plays an important role such as dancing, playing a darts game, or preparing tea at Japanese tea ceremony? We would like to figure it out from the movement of the human joints at an activity, and utilize it to improve our behavior. In our study, a Structure Sensor, Kinect-like 3D image sensor, is used to track and record 3-dimensional positions of the human joints during an action. The device is portable and does not require any sensors nor markers attached to observed persons. We then investigate the effectiveness of a statistical approach. Results of preliminary experiments established for throwing a set of darts will be reported.*

Keywords: Behavior analysis, Human action recognition, 3D image sensor, Kinect, Principal component analysis

1. **Introduction.** In this paper, we try to extract motion features from movement of throwing a set of darts and figure out the key difference in motion between expert darts players and novice ones.

There have been researches to analyze human movements and extract motion features. The most of them aim to classify movements for behavior recognition where it is necessary to remove individuality and extract common features of the movement. Some attempt to extract difference in motion feature to detect special condition such as illness [1-4]. Goals are varied in researches, but target of the analysis is mostly limited to walking behavior.

To track human movement, previous researches use motion capture [5,6], acceleration sensors [2,3,7,8] and/or environmental sensors [9]. Whichever chosen, something like markers or sensors should be attached to human body or environment. We use a Structure Sensor from Occipital, a Microsoft's Kinect-like 3D sensor, which does not require markers nor sensors to be attached to observed players.

Throughout the previous researches, several methods have been proposed for extracting motion features. Main streams are frequency analysis [4,7] and statistical methods [1,8]. We thought the frequency analysis was more suitable for periodic movement like walking and chose PCA (Principal Component Analysis), a well-known statistical analysis method. PCA is mathematically similar to Singular Value Decomposition used in [1] and [8].

There is at least one research to try extracting the motion differences between expert tennis players and novice ones [10]. They used a motion capture to track human motion and extension of DTW (Dynamic Time Warping) [11] to extract motion differences. It is closely related to our research in which we also use DTW but not for extracting motion differences but for its original purpose, to align time series of data.

DOI: 10.24507/icicelb.11.05.487

The structure of the paper is as follows. The procedure of analysis is explained in Section 2. The tools and software used in the experiments are detailed in Section 3. The problem to be solved in this paper is explained in Section 4. The configuration of experiments and results are explained in Sections 5 and 6 respectively. In Section 7, the paper is concluded.

2. **Procedure of Analysis.** Let $\boldsymbol{p}_{ij}(t) = (x_{ij}(t), y_{ij}(t), z_{ij}(t))$ be a 3-dimensional position vector of the j-th joint of the i-th player at the frame number t. A set of joint positions for the i-th player at frame t becomes $\boldsymbol{p}_i(t) = (\boldsymbol{p}_{i1}(t), \boldsymbol{p}_{i2}(t), \ldots, \boldsymbol{p}_{iJ}(t))$ where J is the number of observed joint positions. The obtained sequence of the joint positions for the i-th player becomes $\boldsymbol{p}_i = (\boldsymbol{p}_i(0), \boldsymbol{p}_i(1), \ldots, \boldsymbol{p}_i(T_i))$ where T_i is the total number of observed frames. We here call it a "joint position vector".

We collect the joint position vectors for players and compose a matrix $P = \left(\boldsymbol{p}_1^T, \boldsymbol{p}_2^T, \ldots, \boldsymbol{p}_N^T\right)$ where N is the total numbers of observed players. At this point, we apply FastDTW [12] to adjusting the length of the vectors T_i to the $T = \max_i T_i$ since T_i generally varies for each observation. Furthermore, we normalize P so that the average and the standard deviation of the elements become 0 and 1 respectively. The resulting matrix P can be rewritten as

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1L} & p_{2L} & \cdots & p_{NL} \end{pmatrix}$$

where $L = 3 \times J \times (T + 1)$.

PCA is then applied to the matrix P to compressing and obtaining a principal component matrix Q which can be written as $Q = \left(\boldsymbol{q}_1^T, \boldsymbol{q}_2^T, \ldots, \boldsymbol{q}_M^T\right)$ where M is the number of principal components. $\boldsymbol{q}_i$ is defined by $\boldsymbol{q}_i = (\boldsymbol{q}_{i1}, \boldsymbol{q}_{i2}, \ldots, \boldsymbol{q}_{iL})$ where $\boldsymbol{q}_{ij}$ is the j-th principal component score of the i-th principal component. In general, $\boldsymbol{q}_i$ with the smaller index number i describes the more common features of $\boldsymbol{p}_i$, and contrarily, $\boldsymbol{q}_i$ with the larger index describes the more minor features.

We assume that $\boldsymbol{q}_1$ (and maybe $\boldsymbol{q}_2$) might be similar between expert and novice players when they both perform the same movement. We also assume that the remaining principal component vectors with the larger index numbers describe the minor features which can explain differences in motion between expert and novice players.

3. **Capturing Joint Positions.** A Structure Sensor shown in Figure 1 is used to capture time series of depth images. It is very small with 119.2 mm in width, 28 mm in height, and 29 mm in depth. It can capture 30 or 60 frames per second, where each frame contains spatial information in a range from 40 cm to 3.5 m in depth, 58 degrees in horizontal and 45 degrees in vertical to an accuracy of less than 30 mm.

FIGURE 1. A Structure Sensor mounted on a tripod

At the same time, each of obtained depth images is processed through OpenNI2 and NiTE2 libraries to detect human shape in the image and calculate 3-dimensional positions of the human joints. It can be done almost immediately. Accordingly, for each player, 30 to 60 sets of fifteen joint positions per second can be obtained ($J = 15$ in Section 2). Fifteen positions are head, neck, torso, shoulder (right/left), elbow (right/left), hand (right/left), hip (right/left), knee (right/left), and foot (right/left). A depth image and detected joint positions drawn on the image are shown in Figure 2.

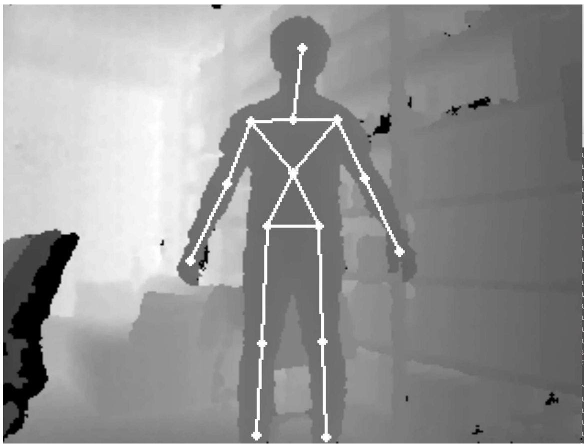


FIGURE 2. An example of captured depth image and joint positions drawn on the image

4. **Target Behavior.** Our target is a motion throwing a set of three darts to a dartboard. We call this a "turn". This is a very basic movement when playing darts games. There are many different darts games but almost all require throwing a set of three darts in a turn.

A "dart" consists of a tip, a metal body called "barrel", and a plastic tail with several solid wings called "flight". In the experiments, we use so-called "soft dart" which is equipped with not a metal tip but a plastic tip.

A 451 mm-wide official size dartboard is fixed on the wall to keep its center at a height of 173 cm from the floor and a distance of 240 cm from the throwing position. The sensor is put under the dartboard at a height of 80 cm from the floor. The configuration of the equipment is shown in Figure 3.

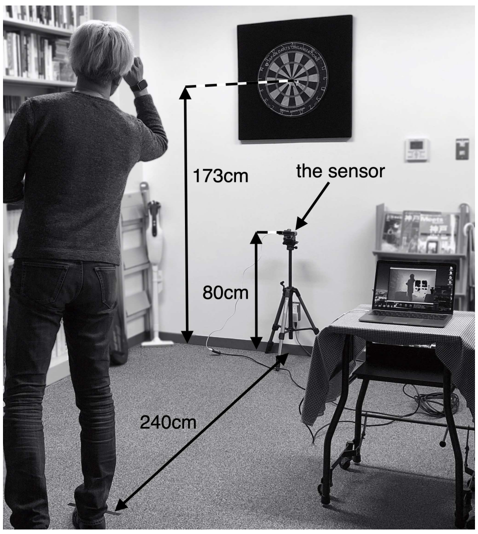


FIGURE 3. The configuration of capturing system

5. **Experiment.** We asked 7 participants to throw a set of darts for several turns. We totally obtained joint trajectories for 37 turns. Since throwing paces are different individually, the time taken for a turn varied from about 3 to 10 seconds. In the experiment, we configure the sensor with 30 frames per second, and consequently, we could capture about 100 to 300 frames per player. We actually could capture 109 frames at the minimum and 317 frames at the maximum ($T = 317$ in Section 2).

Also, the players were classified into expert, intermediate, and novice players according to their experience. As a result, 2 players were distinguished as expert, 2 players intermediate, and remaining 3 players as novice. Hereafter, (ex), (mid), and (nb) at the index indicate expert, intermediate and novice respectively. As a result, we could obtain $[14,265 \times 37]$ matrix as P.

6. **Result.** We applied PCA to the obtained matrix P and got totally 23 principal components ($M = 23$ in Section 2). The contribution ratios of the first 6 of them and cumulative ones are shown in Table 1 where PC stands for principal component. According to the cumulative contribution ratios, the first 4 principal components can explain about two thirds of movements. It is not enough large but the remaining 19 principal components only have very small contribution ratio of less than 5.4% and we decided to investigate them.

Table 1. Contribution ratios of principal components

	PC1	PC2	PC3	PC4	PC5	PC6
Ratio (%)	25.5	17.7	13.1	7.0	5.4	4.2
Cum. (%)	25.5	43.2	56.3	63.3	68.7	72.9

Observed trajectories of joints in the experiment are mapped onto the principal components space by using the following equation:

$$\mathrm{PC}_{ij} = \boldsymbol{q}_i \cdot \boldsymbol{p}_j^T$$

where i is the index of the principal components and j is the index of observed trajectory. They are plotted in Figure 4 where turns by the same players are plotted with the same markers. The color of markers denotes the ability of the players: black markers for expert players 'A1' and 'A2', gray markers for intermediate 'B1' and 'B2', and white markers for novice 'C1', 'C2' and 'C3'. As shown in the figure, the expert, the intermediate, and the novice are well separated in the plot of PC2-PC3. It means that these two principal components of PC2 and PC3 seem the most important to distinguish the player levels between expert, intermediate and novice.

Figure 5 shows $\boldsymbol{q}_i$ for $i = 1, 2, 3, 4$, the component scores of PC1 to PC4. Figure 5(a) plots an average score over joints, and Figure 5(b) plots an average score over frames. In both plots, the darker pixels correspond to greater values and the brighter to smaller values.

There are three darker areas followed by lighter area seen around 50, 150, and 250 frames in Figure 5(a). This oscillation of $\boldsymbol{q}_i$ is corresponding to the movements throwing three darts. It is rather clearly observed in PC4. However, no significant characteristics can be observed in PC2 and PC3.

In Figure 5(b), it is quite interesting that both PC1 and PC2 have the larger values for the left hand and the left foot. It is understandable that values for the right hand and the right elbow become large in PC1 since all the players threw darts by the right arm. However, the left hand only holds remaining darts and the left foot is free from the weight during throwing. Values for the left elbow, the left shoulder, the left knee and the left hip are also large in PC2. Neck and head also have the large values in PC3. These results hint that the movement of the left half of the body and the head is very important

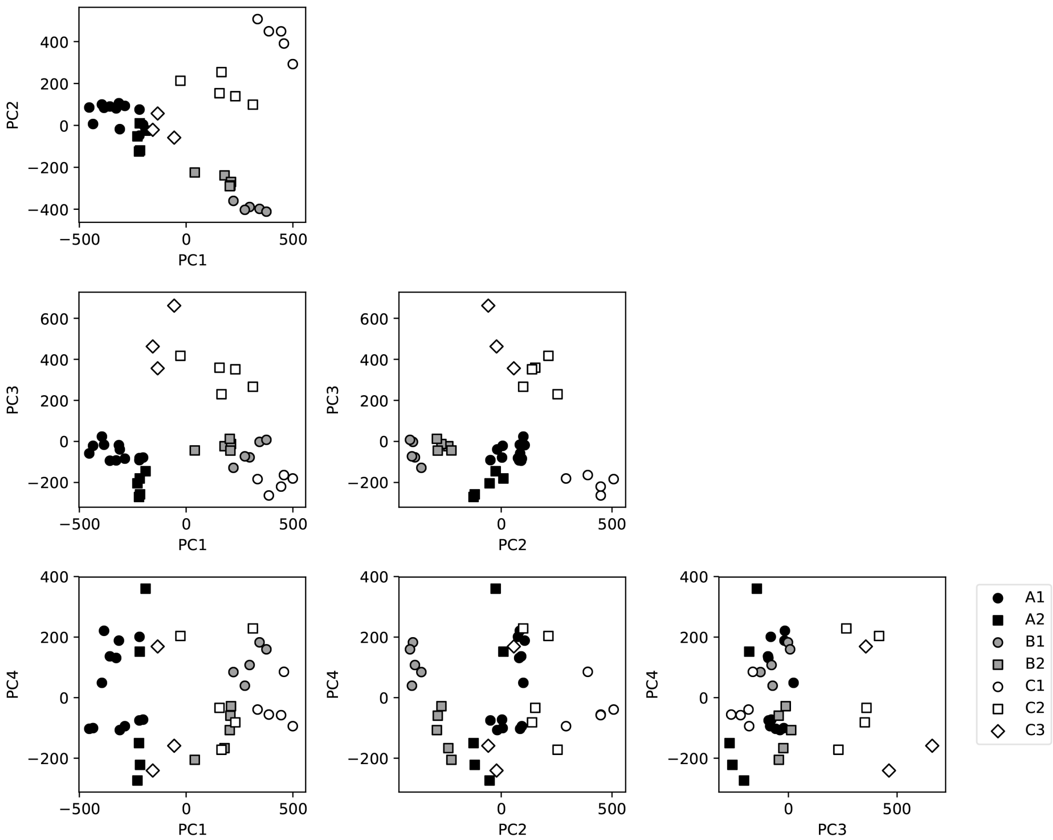
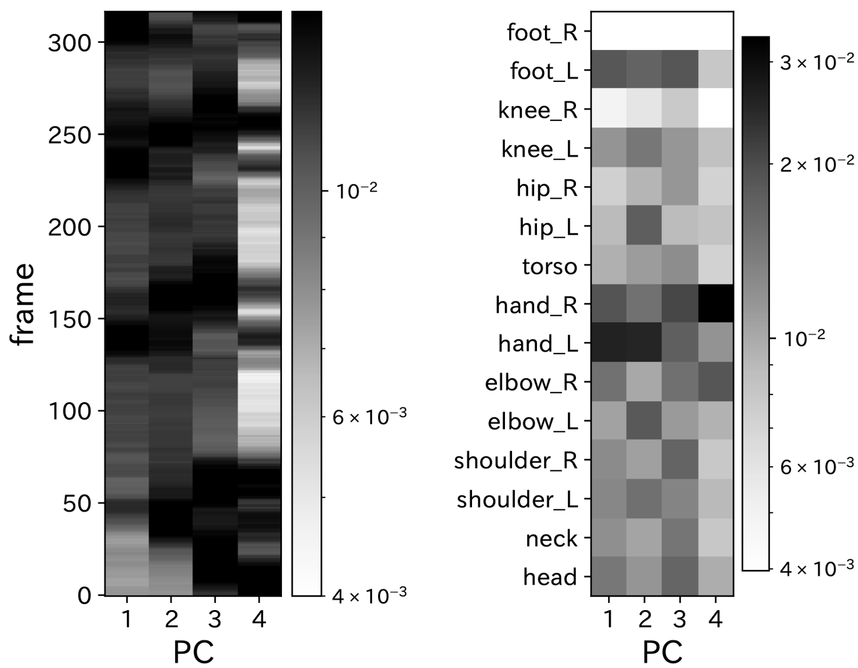


Figure 4. Movements throwing darts mapped on the principal components space



(a) Sorted by the frame (b) Sorted by the joint

Figure 5. A plot of $\boldsymbol{q}_i$ for $i = 1, 2, 3, 4$

for playing darts well, though throwing darts by the right arm and the body weight is applied to the right leg.

7. **Conclusion.** In this study, we have tried to use Kinect-like 3D sensor and PCA to the behavior analysis to find the key movements distinguishing expert and novice players. It is applied to darts-throwing and obtained an interesting result of suggesting importance of the movement of the left half of the body. It is usually thought not important much. Further development of analysis method is required for investigating key movements but we believe we could show the possibility of the approach.

Acknowledgment. This research is partially supported by JSPS Grant-in-Aid for Scientific Research (C), Grant Number 19K12247. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] K. Mishima, S. Kanata, H. Nakanishi, T. Sawaragi and Y. Horiguchi, Extraction of similarities and differences in human behavior using singular value decomposition, *Journal of the Institute of Electronics, Information and Communication Engineers A*, vol.J94-A, no.4, pp.292-302, 2011 (in Japanese).

[2] M. Sekine, T. Tamura, M. Yoshida, Y. Suda, Y. Kimura, H. Miyoshi, Y. Kijima, Y. Higashi and T. Fujimoto, A gait abnormality measure based on root mean square of trunk acceleration, *Journal of NeuroEngineering and Rehabilitation*, vol.10, no.1, p.118, 2013.

[3] A. Mannini, D. Trojaniello, A. Cereatti and A. Sabatini, A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and Huntington's disease patients, *Sensors*, vol.16, no.2, p.134, 2016.

[4] S. Nakajima, M. Ikebuchi and T. Sugawara, A method to distinguish between patients with hemiplegic stroke and others using linear distribution of frequency analysis of acceleration of gait and back propagation, *The 12th International Conference on Innovative Computing, Information and Control*, 2017.

[5] B. Zhang, S. Jiang, K. Yan and D. Wei, Human walking analysis, evaluation and classification based on motion capture system, in *Health Management – Different Approaches and Solutions*, K. Smigorski (ed.), InTech, 2011.

[6] J. Lei, Q. Zhu, J. Wu and R. Xiong, Gait analysis of human locomotion based on motion capture system, *2014 the 9th IEEE Conference on Industrial Electronics and Applications*, pp.765-769, 2014.

[7] N. Wang, E. Ambikairajah, N. H. Lovell and B. G. Celler, Accelerometry based classification of walking patterns using time-frequency analysis, *2007 the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.4899-4902, 2007.

[8] T. Akiduki, A. Uchida, Z. Zhang, T. Imamura and H. Takahashi, Extraction of human gait feature from acceleration data, *ICIC Express Letters, Part B: Applications*, vol.7, no.3, pp.649-656, 2016.

[9] A. Muro-de-la Herran, B. Garcia-Zapirain and A. Mendez-Zorrilla, Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications, *Sensors*, vol.14, no.2, pp.3362-3394, 2014.

[10] M. Morel, R. Kulpa, A. Sorel, C. Achard and S. Dubuisson, Automatic and generic evaluation of spatial and temporal errors in sport motions, *The 11th International Conference on Computer Vision Theory and Applications (VISAPP 2016)*, pp.542-551, 2016.

[11] H. Sakoe and S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.26, no.1, pp.43-49, 1978.

[12] S. Salvador and P. Chan, Toward accurate dynamic time warping in linear time and space, *Intell. Data Anal.*, 2007.