The possibility of personality extraction using skeletal information in hip-hop dance by human or machine

Saeka Furuichi[1], Kazuki Abe[1], Satoshi Nakamura[1]

[1] Meiji University, Nakano 4-21-1 Nakano-ku, Tokyo, Japan
`ev60533@meiji.ac.jp`

Abstract. The same dance can give different impressions depending on the way the dancers convey their own emotions and personality through their interpretation of the dance. Beginner dancers who are teaching themselves often search for dance videos online that match their own personality in order to practice and mimic them, but it is not easy to find a dance that suits their own personality and skill level. In this work, we examined hip-hop dance to determine whether it is possible to identify one's own dance from skeleton information acquired by Kinect and whether it is possible to mechanically extract information representing the individuality of dance. Experimental results showed that rich experienced dancers could distinguish their own dances by only skeleton information, and it was also possible to distinguish from averaged skeletal information. Furthermore, we generated features from the skeletal information of dance and clarified that individual dance can be distinguished accurately by machine learning.

Keywords: Dance, Personality, Kinect, Skeleton, Random forest.

1 Introduction

In Japan, interest in dance has been increasing, and dance has become mandatory in physical education classes in junior high school since 2012 as a means of improving the ability to express oneself and communicate. In addition, on video sharing websites such as *Nico Nico Douga*, dance videos uploaded by individuals are extremely popular. As the number of people who enjoy dancing increases, there has also been a huge demand for learning to dance well.

One way to improve dancing ability is to go to a dance class and learn from a skilled instructor. However, it has also become common for beginners to watch dance videos online and teach themselves, as a large number of dance videos are available on the Web. The advantages of learning from an instructor are that users can learn and practice choreography that matches their level and they can receive corrections or advice for elements to work on. In contrast, when teaching themselves, they need to find appropriate choreographies themselves, which can be hard due to the difficulty of objectively grasping their own level. It is also somewhat complicated to search for dance videos at the right level from among the overwhelming number of videos on such sites.

Here, personality is one of the key elements of dance. In dance, personality leads to broadening the expression of dance and add uniqueness to one's own dance style. By developing personality, it is possible to dance in a way that is more attractive to the audience. In contrast, when personality is not clearly expressed in dance, even if the dancers have sufficient skill, their performances tend to be superficial and boring. Therefore, it is crucial that dancers develop their own personality.

Although many studies [1] [2] have focused on dance level, very few have focused on the personality of dance, and the specific elements of dance in which personality appears are not yet known. In our past work [3], we investigated whether dancers could distinguish their own dance based only on skeletal information on their movement. The results showed that dancers could subjectively determine their own dance to a certain extent, but the specific movements or features in which personality appeared were not clear. To investigate the personality, we have to clarify which features work most effectively when the system judges the personality mechanically by machine learning.

One way to observe the dancing movement of a person is to use skeletal information of the human body obtained by techniques such as motion capture. In this paper, we examine hip-hop dance to identify whether it is possible to extract personality in dance, specifically, by using approximate skeletal information obtained not from high-precision motion capture but from inexpensive depth cameras such as Kinect and methods that can obtain skeletal information from moving images such as OpenPose [4]. We use only skeletal information and avoid information that is not related to actual dancing, such as physique and appearance. Then, we conduct the experimental test by human and by machine learning techniques, and clarify the possibility to extract personality.

2 Related Work

To extract motion characteristics for hip-hop dance evaluation, Sato et al. [5] focused on hand waves and clarified that the constant propagation speed is the most important factor to feel smooth waving. Chan et al. [1] proposed a method of making an average dancer appear better by using images of the movements of a more skilled dancer. These studies have extensively analyzed the movements and features of dance, but the characteristics and personality of an individual's dancing style remain unclear. In this research, we investigate which elements of dance are related to the personality of dance.

There were several studies on dance education. Yonezawa et al. [6] studied changes in the attitudes of elementary and junior high school teachers who began to add dance to the curriculum, as well as the effects of this curriculum. Yamaguchi et al. [7] proposed a system to support dance education by generating sounds in real time according to the dance, thus supporting the creativity of beginner dancers. Nakamura et al. [8] demonstrated a device that uses vibration to tell the user when to start a dance action, and Yang et al. [9] used VR to show beginners the movements of dance experts and have conducted research to support dance improvement by imitating them. Fujimoto et al. [2] proposed a method for beginners to identify and improve their dance form by mapping their movements to those of more experienced dancers. The purpose of these studies is to support the creativity of dance, to improve individual movements, and to

mimic the movements of dance experts. In addition to these studies, it is considered that clarifying the personality of dance will be helpful for improving dance skills and helping beginners make rich expressions that capture their personality. Therefore, in this research, we investigate the personality of dance and use it as a foothold to support dance advancement with personality.

Studies on dance videos have also been increasing as the release of individual dance videos has become more popular. Tsuchida et al. [10] proposed an interactive editing system for multi-view dance videos that can be used to easily create attractive dance videos without requiring video editing expertise. They also developed a new searching system [11] that utilizes the user's dance moves as a query to search for videos that include music appropriate for that style of dancing. However, the system does not take personality into consideration, as it is characterized by whether the choreography itself is similar. The purpose of our study is to clarify how these points are identified in order to consider personality.

On the other hand, researches using body motion data are widely conducted. Mousas [12] studied the structure of dance motions by hidden Markov model (HMMs) and developed a method to make dance motions natural on VR. Aristidou et al. [13] conducted a study on dance motions and emotions using Laban Movement Analysis (LMA) and classified dance motions related to emotions. In addition, Senecal et al. [14] analyzed behavior that expresses the emotion of performers and proposed a system for emotional behavior recognition. These studies investigate human movement, but do not take the performer's personality into account. The purpose of this research is to focus on the personality of the dance movement and to clarify whether there is a difference among people.

3   Dataset Construction for Dance Skeleton

In this research, we investigate whether information that expresses the personality of dance can be extracted from the skeletal information of dance. We worked with participants who have dance experience and constructed the skeletal information (dance skeleton dataset) of actual dance.

Skeletal information of the human body is extracted from dancing movement by using Kinect, a motion sensor device. The participants were 22 university students (seven males, 15 females) in a university dance club with dance experience ranging from five months to six years (average: 2.4 years). We asked them to dance about 15 seconds (seven bars) of specific choreography five times and used Kinect to extract skeletal information. The information obtained from Kinect was composed of the 15-point 3D coordinates shown in Fig. 1 (left).

The choreography used to construct the dataset is classified as hip-hop dance and is characterized by a large number of movements featuring the entire body, including raising the legs, squatting, turning, and hitting the chest. The participants practiced this choreography on a daily basis. The music used was Traila$ong's "Gravity".

We used a large space for the experiment and the participants danced in front of the Kinect. These participants all belong to the same dance club at Meiji University and

had 1-hour practice sessions twice a week for three weeks, so their dance performance during the data collection was sufficient. Each participant danced five times and then answered a brief questionnaire about their dance experience. There was a short break between each dance performance. In total, the dataset consisted of 110 items of data comprising five 15-sec dance samples for each of the 22 participants.
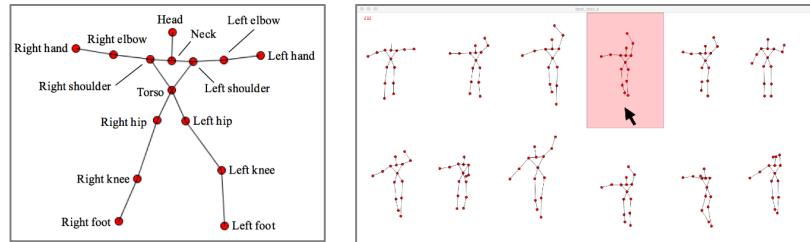

Fig. 1. (left) Fifteen points of skeletal information captured by Kinect. (right) Screenshot of the experiment system.

4　Dance Personality Estimation by Subjective Evaluation

In this paper, we investigate whether it is possible to distinguish one's own dance from the video of only the skeletal information. In addition, to clarify where the personality of the dance appears in the skeletal information, we administered interviews to determine the distinction. Here, participants were divided into two groups: 11 with little dance experience (average 1.0 year) and 11 with relatively rich dance experience (average 3.6 years). This was done because the manner of expressing personality differs depending on the level of dance experience.

In the experiment, as shown in Fig. 1 (right), we prepared a task in which participants were asked to select which dance they felt was their own from among dance images of several people presented as skeletal information only. Twelve different dance images were presented: 11 performed by the participants in that group and one from a member of the other group who had the most similar amount of dance experience. This was done to increase the number of participants which close to the dance level. For the selection of dance, participants were asked to rank the skeletons that they thought were similar to their own dance style from first to third place. The experiment was conducted five times for each item of skeletal information and average skeletal information, and the position of the skeleton dance on the screen was randomized each time. Scores were determined in accordance with the order provided by the participants, and the average score of each skeleton was considered. The score was 5 points for first place, 3 points for second place, and 1 point for third place.

The experimental result showed that among the participants with little dance experience, two out of the 11 participants gave their own dance of skeletal information the highest score and one out of 11 participants gave their own dance of average skeletal information the highest score. In addition, while there were few participants who could accurately identify their own dance, those who did not choose their own dance tended

to choose one dancer (i.e., the same one) exclusively. On the other hand, among the participants with rich dance experience, five out of the 11 participants gave their own dance of skeletal information the highest score and six of the 11 gave their own dance of average skeletal information the highest score. This indicates that, compared to the dancers with little experience, these participants were able to identify their own dance with greater accuracy. Also, here too, participants who did not choose their own dance tended to choose one person exclusively.

After the experiments, we conducted an interview to precisely see which spot in the skeletal information was judged to be similar to one's own dance style. From the interview, we found that participants with rich experience tended to respond by designating the factors in detail. For example, seven out of the 11 participants with little experience stated simply that they moved specific body parts (e.g., arms, knees, and whole body), and six mentioned the shape of the hand (e.g., position and bending condition). In contrast, among the participants with rich experience, ten of the 11 made comments on the hand shape (e.g., the position and angle of the hand), and most of them explained they felt that their personality was likely to appear in the hand position. Six of the 11 stated that they used speed for parts on the outer side of the body (e.g., hands and feet) and whole body. Moreover, four of the five participants who had correctly identified their own dance in the skeletal information experiment commented about the characteristics of their own dance style.

The participants with rich experience provided more detail about their selection process and had a higher selection accuracy. From this, we can assume that they understood the features of their own dances more firmly. In particular, ten of the 11 mentioned hand shape, which indicates that their personality tends to appear in hand. Moreover, none of them mentioned movement, which was the most frequent response among the less experienced participants, so we can conclude that these experienced participants place more emphasis on individual form than on movement style.

5 Dance Personal Estimation with Machine Learning

In order to clarify whether the skeletal information can convey the personality of the dancer, we analyzed whether it is possible to distinguish the individuals with machine learning from the skeletal information obtained using Kinect. Here, we conducted an experiment to classify the dance for each individual by generating feature quantities based on the discrimination factors used by the participants obtained from the interview in the subjective evaluation experiment.

According to the responses to the interview in the subjective evaluation experiments, many participants relied on the shapes of their hands and feet to identify their own dance from others'. Therefore, for the dance in the constructed dataset, we calculated the joint angles of six places of the elbow, the shoulder, and the knee on the left and right for each frame based on the 3D skeletal coordinates of 15 points obtained by Kinect. Next, we generated a 6D vector by calculating the average of angles every second (30 frames). Here, since the dance samples in the dataset are less than 15 seconds long, these 6D vectors can be acquired for up to 14 seconds. We combined 84 dimensions of

14 sec × 6 dimensions into one and generated them as angle features. At this time, as a countermeasure for noise, we did two approaches, linear interpolation from the frame before and after the defect frame, and smoothing by the method of exponential moving average.

We also generated feature values for motions in which the response was high among participants with little experience. Here, with regard to the 3D skeletal information of the chest and the 13 points of the left and right hands, elbows, hips, knees, and feet, the amount of spatial movement of each skeletal point in one second was acquired as a 13D vector. As with the angular features, the dance samples in the dataset are less than 15 seconds long, so these 13D vectors can be obtained for up to 14 seconds. We combined 182 dimensions of 14 seconds × 13 dimensions into one and generated them as movement features. For these two features, we compare the estimation accuracy when each feature is used.

We divided participants into two groups—those with little dance experience (the same individuals as in the experiment in section 4) and those with rich experience—to determine how well dancers can be identified from skeletal information.

We used a random forest as the classifier algorithm. Three dances out of five were used as training data and the remaining two as test data. Each group includes 60 data of 12 persons × 5 times, and for the two dances used as test data, 36 data are training data, 24 data are test data, and 12-value classification learning is performed. In addition, since $_5C_2 = 10$ combinations of training and test data can be made from data per person, we learned ten patterns each time and calculated the average accuracy rate and classification probability from the results.

	a's	b's	c's	d's	e's	f's	g's	h's	i's	j's	k's	l's
a	0.25	0.07	0.10	0.05	0.05	0.09	0.08	0.07	0.05	0.09	0.05	0.06
b	0.05	0.40	0.06	0.04	0.07	0.07	0.06	0.07	0.02	0.06	0.04	0.06
c	0.08	0.07	0.27	0.03	0.04	0.08	0.07	0.07	0.06	0.09	0.06	0.07
d	0.06	0.06	0.04	0.31	0.08	0.06	0.06	0.08	0.08	0.07	0.06	0.03
e	0.05	0.08	0.04	0.07	0.26	0.08	0.09	0.06	0.05	0.08	0.09	0.05
f	0.07	0.06	0.06	0.04	0.06	0.24	0.09	0.10	0.04	0.09	0.08	0.06
g	0.06	0.06	0.05	0.06	0.08	0.09	0.30	0.07	0.05	0.07	0.07	0.07
h	0.09	0.07	0.06	0.05	0.05	0.10	0.07	0.26	0.06	0.08	0.08	0.09
i	0.05	0.03	0.07	0.07	0.05	0.07	0.07	0.07	0.28	0.11	0.09	0.04
j	0.06	0.05	0.08	0.04	0.06	0.09	0.06	0.09	0.07	0.26	0.08	0.06
k	0.03	0.03	0.05	0.04	0.07	0.08	0.06	0.07	0.05	0.07	0.38	0.06
l	0.04	0.06	0.06	0.02	0.05	0.05	0.07	0.10	0.03	0.07	0.08	0.36

	A's	B's	C's	D's	E's	F's	G's	H's	I's	J's	K's	L's
A	0.26	0.04	0.04	0.14	0.09	0.05	0.04	0.08	0.09	0.08	0.05	0.03
B	0.03	0.37	0.07	0.04	0.04	0.07	0.06	0.05	0.05	0.06	0.09	0.06
C	0.04	0.07	0.32	0.04	0.04	0.09	0.03	0.06	0.10	0.08	0.08	0.03
D	0.14	0.04	0.05	0.24	0.10	0.05	0.05	0.08	0.10	0.08	0.05	0.03
E	0.08	0.04	0.04	0.09	0.37	0.08	0.03	0.07	0.08	0.07	0.04	0.02
F	0.05	0.07	0.05	0.05	0.08	0.34	0.03	0.05	0.08	0.09	0.09	0.03
G	0.04	0.07	0.09	0.04	0.03	0.05	0.24	0.06	0.06	0.09	0.09	0.15
H	0.07	0.07	0.03	0.09	0.06	0.05	0.06	0.26	0.10	0.08	0.06	0.03
I	0.08	0.05	0.06	0.09	0.08	0.07	0.05	0.11	0.18	0.10	0.08	0.05
J	0.08	0.05	0.08	0.07	0.07	0.08	0.07	0.07	0.09	0.20	0.08	0.06
K	0.05	0.09	0.08	0.04	0.04	0.08	0.07	0.06	0.08	0.09	0.25	0.08
L	0.03	0.08	0.10	0.03	0.03	0.04	0.16	0.05	0.06	0.08	0.10	0.25

Fig. 2. (left) The average classification probability of learning using angle features for participants with little experience. (right) The average classification probability of learning using angle features for participants with rich experience.

Fig. 2 shows the classification probabilities when learning is performed using angle features. The columns in the table indicate the participants to be classified and the rows show each dance. For example, in the left table, when "a" in the column and "a's (dance)" in the row are selected, the classifier estimated the dance of participant "a", then it is determined that participant "a" has a probability of 0.25 on average. The total of the table row direction is 1.0. The classifier outputs the most probable of these classification probabilities as the estimation result. The stronger the background color is, the higher the classification probability of each participant is. According to the tables,

in both groups, each participant's own dance had the highest classification probability. In addition, the average accuracy rate for ten iterations of learning was high: 99.1% for the group with little experience and 92.0% for the group with rich experience. From these results, we conclude that it is possible to distinguish individuals by machine learning, with joint angles as feature quantities. In addition, similar results were obtained with other classifications used movement amount of the skeleton point.

Table 1. The average classification accuracy by machine learning in each feature.

	Little experience	Rich experience
Angle feature	99.1%	92.0%
Movement feature	95.4%	89.5%

Table 1 shows the average classification accuracy in each experience group and by using each feature. From these results, we conclude that it is possible to distinguish individuals using machine learning, with joint angles as feature quantities.

Here, the random forest is ensemble learning that learns by a set of decision trees, and the importance of each feature vector can be evaluated by comparing each decision tree. We therefore measured which feature vector was effective for the features in each learning. Regarding angle features, we found that the angle of the left and right knees is a feature vector with relatively high importance for both groups. In addition, regarding the movement feature, the importance of the feature vectors in the upper body (such as the chest and right shoulder) tended to be high for both groups.

6 Discussion

Here, we discuss the personality of dance on the basis of the results of the subjective evaluation in Chapter 4 and the results of the machine learning in Chapter 5.

In the subjective evaluation, we observed a tendency to continuously select the movement of a specific participant, and it became clear that it was possible to find features subjectively from dances with only skeletal information. However, not many of the participants could correctly identify their own dance. Of course, it can be difficult to objectively grasp one's own dance in ordinary practice, even if you have rich experience in dance. On the other hand, in the experiment by machine learning, although learning was performed using two features (angle and movement amount), it was possible to discriminate dances with high accuracy regardless of the feature used. We conclude that there are many dancers who cannot recognize their own personality despite the existence of characteristics and movements that point to this personality, and that it should become easier to find the personality of one's own dance by using these characteristic features.

Next, we consider the position and elements of the body where personality tends to appear. In the subjective evaluation, most of the participants concentrated on the hands in the skeleton regardless of experience, and more than half judged the balance of the whole body. In addition, it became clear that the participants with less experience placed more emphasis on movement and the participants with more experience placed

emphasis on the shape. On the other hand, in the experiment by machine learning, the learning by angle features emphasized the shape of the foot, and in the learning by movement amount features, the movement of the chest was emphasized. Two factors—the degree of bending of the foot and the movement of the chest (upper body)—are considered to correspond to the balance of the whole body emphasized in the subjective evaluation, because the ratio occupied in the video used for the subjective evaluation is large. On the other hand, with regard to the hand having the highest number of mentions from participants in the subjective evaluation, which was not emphasized in the experiments using machine learning, there is a possibility that personality appears in the direction and shape of their hand rather than its position. Therefore, by focusing on the state of hands, it is possible that the individuality can be further highlighted.

In order to make use of the results of this study, there are several challenges to be overcome in exploring dance videos for learning independently. In this research, we asked participants to dance to specific choreography, and we performed experiments from two points of view, namely, subjective evaluation and machine learning. However, when we actually explore dance videos, the dancers use different choreographies for each video. Therefore, it is necessary not only to be able to identify the person by comparing dances with the same choreography but also to be able to determine the individuality of the person in completely different choreography. It is therefore necessary to conduct additional experiments with a greater variety of choreography, and to clarify whether it is possible to identify the dancer in such cases. We also need to acquire personality and implement a mechanism to search for dance videos based on this.

7 Conclusion

In this paper, we investigated hip-hop dance to determine whether it is possible to extract personality in dance from skeletal information acquired by Kinect. In a subjective evaluation, we found that more experienced participants could distinguish their own dance from only the skeletal information, and they could also be judged from the average skeleton as well. We also found that the main points for judging dance differed depending on the level of experience: for example, less experienced participants tended to emphasize the way of movement and more experienced ones emphasized the hand shape. In dance estimation by machine learning using skeletal features, we found through learning and comparing with two features, namely, angle and movement amount, that it is possible to discriminate individuals with high accuracy using either of them. Overall, using the angle feature was more accurate, as it is close to the judgment criteria of the experienced participants.

In the future, we intend to clarify what constitutes personality in dance by extending these analyses and to further examine the application method of the extracted personality. We will also consider a method to search for dance videos that have completely different choreography but that match the personality of the user.

References

1. Chan, C., Ginosar, S., Zhou, T., Efros, A., A.: Everybody Dance Now. In: arXiv 2018, vol. 1, no. 1 (2018).
2. Fujimoto, M., Tsukamoto, M., Terada, T.: A Dance Training System that Maps Self-Images onto an Instruction Video. In: The Fifth International Conference on Advances in Computer-Human Interactions (ACHI2012), pp. 309-314. Valencia (2012).
3. Saito, H., Maki, Y., Tsuchiya, S., Nakamura, S.: Can People Sense Their Personalities Only by Watching the Movements of Their Skeleton in Street Dancing Performances? In: Kurosu M. (eds) Human-Computer Interaction. Interaction Technologies (HCI 2018), Lecture Notes in Computer Science, vol 10903. Springer, Cham (2018).
4. Cao Z, Simon T, Wei, S.E., and Sheikh Y.: Real time multi-person 2D pose estimation using part affinity fields. In: Computer Vision and Pattern Recognition(CVPR2017), Honolulu (2017).
5. Sato, N., Imura, S., Nunome, H., Ikegami, Y.: Motion characteristics in hip hop dance underlying subjective evaluation of the performance. In: The 30th Conference of the International Society of Biomechanics in Sports (ISBS2012), pp. 17-20, Melbourne (2012).
6. Yonezawa, M.: The investigations on the teachers' attitudes to dance in the face of scholastic requirement of dance in middle schools in Heisei 24 (2012) academic year. In: Studies in Humanities of Kanagawa University, vol.178, pp. 53-80. Kanagawa (2012).
7. Yamaguchi, T., Kadone, H.: Supporting creative dance performance using a grasping-type musical interface. In: 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO2014), Bali (2014).
8. Nakamura, A., Tabata, S., Ueda, T., Kiyofuji, S., Kuno, Y.: Dance training system with active vibro-devices and a mobile image display. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), Edmonton (2005).
9. Yang, U., Kim, G.: Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System. In: Presence, volume 11, issue 3 (2002).
10. Tsuchida, S., Fukayama, S., Goto, M.: Automatic System for Editing Dance Videos Recorded Using Multiple Cameras. In: Cheok A., Inami M., Romão T. (eds) Advances in Computer Entertainment Technology. (ACE2017). Lecture Notes in Computer Science, vol 10714. Springer, Cham (2017).
11. Tsuchida, S., Fukayama, S., Goto, M.: Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity, Proc. of the 25h International Conference on Multimedia Modeling (MMM2019), pp. 251—263, Thessaloniki (2019).
12. Mousas, C.: Performance-Driven Dance Motion Control of a Virtual Partner Character. 57-64 (2018).
13. Aristidou, A., Charalambous, P., Chrysanthou, Y.: Emotion Analysis and Classification: Understanding the Performers' Emotions Using the LMA Entities. In: Computer Graphics Forum. vol. 34, pp. 262–276 (2015).
14. Senecal, S., Cuel, L., Aristidou, A., Thalmann, N.: Continuous body emotion recognition system during theater performances. In: Computer Animation and Virtual Worlds. vol. 27, pp. 311-320 (2016).