

Representation of Fundamental Movements and Pauses for Archiving Traditional Skills

Hiroki Nomiya and Teruhisa Hochin

Abstract Considering the reduction of the number of persons engaging in traditional skills, archiving various types of traditional skills is required to preserve and transmit them to future generations. We focus on representing fundamental movements and pauses in traditional skills because they are key components in describing traditional skills and archiving them. The fundamental movements and pauses can be described based on the movements of a number of body parts obtained using motion capture system. In this paper, we propose an efficient method to represent fundamental movements and pauses using the motion data. The proposed method generates concise and informative feature values from the motion data on the basis of dimensionality reduction and feature selection. The effectiveness of the proposed method is evaluated through an experiment to describe several types of fundamental movements in Japanese traditional tea ceremony.

Keywords Archiving · Traditional skill · Fundamental movement · Pause · Dimensionality reduction · Feature selection

1 Introduction

There are a wide variety of valuable traditional crafts and industries. It becomes, however, more and more difficult to preserve and transmit them because young people are less interested in traditional cultures and not willing to obtain traditional skills. Moreover, aging of the skilled people makes it more difficult to transmit the traditional skills [1–3]. Therefore, preserving traditional skills is urgently needed.

Most of the traditional skills and industries are founded on various traditional skills. It will be thus effective to archive traditional skills for preserving and

H. Nomiya (✉) · T. Hochin

Kyoto Institute of Technology, Goshokaido-cho, Matsugasaki, Sakyo-ku, Kyoto, Japan
e-mail: nomiya@kit.ac.jp

T. Hochin

e-mail: hochin@kit.ac.jp

transmitting them. The archives will be helpful for the people interested in traditional cultures.

In order to archive traditional skills, representing them as a certain type of digital data is important since such type of data can be easily and widely utilized by commonly-used personal computers. For example, video data can be used as an archive. It is relatively easy to make an archive by recording traditional skills using a video camera. However, the amount of video data tends to be very large and thus archiving a wide variety of traditional skills requires a large-scale storage system. In addition, it is difficult for novices to obtain traditional skills by simply watch the video.

In this paper, we propose a method to concisely represent traditional skills using motion capture system. The system captures the movements of several body parts of a skilled person using a number of markers attached on the body parts. The movement is represented by time-series data of the position of each marker. We define some feature values on the basis of the motion data. The amount of raw motion data is very large due to the high sampling frequency of the motion capture system. We intend to fully reduce the amount of data by applying dimensionality reduction to the raw motion data. Additionally, we introduce a feature selection method to further reduce the amount of data and find out the important movement. Our feature values can be defined for each marker. Thus, selected features indicate the important movement of a certain body part. This will be helpful to obtain traditional skills.

We attempt to represent a traditional skill by decomposing it into several fundamental movements because a traditional skill generally consists of several fundamental movements. In this paper, we focus on Japanese traditional tea ceremony as one of the representative traditional skills in Japan. It includes a sequence of fundamental movements. Additionally, a pause, which is called “Ma” in Japanese, is considered to be important as well as the fundamental movements [4]. We conduct an experiment to represent seven types of fundamental movements and the pause in Japanese traditional tea ceremony. The proposed method is evaluated from the viewpoints of the conciseness and accuracy of the representation.

The remaining of this paper is organized as follows. Section 2 presents related work. Section 3 explains the representation of fundamental movements and pauses. Section 4 evaluates the representation through an experiment. Section 5 gives a consideration about the experimental result. Finally, Sect. 6 concludes this paper.

2 Related Work

The movements of traditional skills have been studied to transmit them for future generations. For example, the movements of craft works have been analyzed [1–3]. The pauses in traditional skills have been investigated as well [4]. A learning system has been provided by visualizing the movements of the skills [5]. It is, however, still difficult to precisely represent fundamental movements and pauses of various traditional skills due to their complexity.

The motion data could be effectively utilized to analyze and describe fundamental movements in traditional skills [4]. For the analysis of the motion data, dynamic time warping [6] or Fourier transform [4] have been widely used. However, such methods have a problem of high computational cost.

We have proposed a method to concisely represent fundamental movements and pauses by introducing dimensionality reduction into motion data on the basis of a quantization technique [7]. The amount of data can be considerably reduced by this method. However, this method simply reduces the motion data of each marker. Hence, it is impossible to distinguish important markers from unimportant ones.

In this paper, we make it possible to find out important markers as well as reduce the amount of data. This will be useful for obtaining traditional skills as well as developing concise archive systems.

3 Representation of Fundamental Movements and Pauses

3.1 Raw Motion Data

The motion capture system we use provides the coordinates of twenty-nine body parts in the three-dimensional Euclidean space. Raw motion data can be represented as the sequences (i.e., time series) of the x-, y-, and z-coordinates of the body parts. This is the same system used in our previous work [7]. The motion M_i of the i th marker is given as the time series represented by Eq. (1).

$$M_i = (x_{i,1}, y_{i,1}, z_{i,1}), \dots, (x_{i,N}, y_{i,N}, z_{i,N}). \quad (1)$$

In this equation, N is the length of the time series (we call N number of frames), and $x_{i,n}$ ($y_{i,n}$, and $z_{i,n}$, respectively) is the x-coordinate (y- and z-coordinates) of the i th marker at the n th frame.

The markers used to record the coordinates are shown in Fig. 1. Figure 1a shows a skilled person with the markers and (b) illustrates the marker numbers.

3.2 Dimensionality Reduction

The aforementioned motion capture system records the positions of markers 100 times per second. Hence, 8700 real values are generated for each second since there are 29 markers and the position of each marker represented by 3 real values (i.e., x-, y-, and z-coordinates). Using all the values leads to high computational complexity. For the purpose of efficient computation, we apply dimensionality reduction to raw motion data.

At first, the x-, y-, and z-coordinates in the Cartesian coordinate system are converted to the radius, inclination, and azimuth in the spherical coordinate system

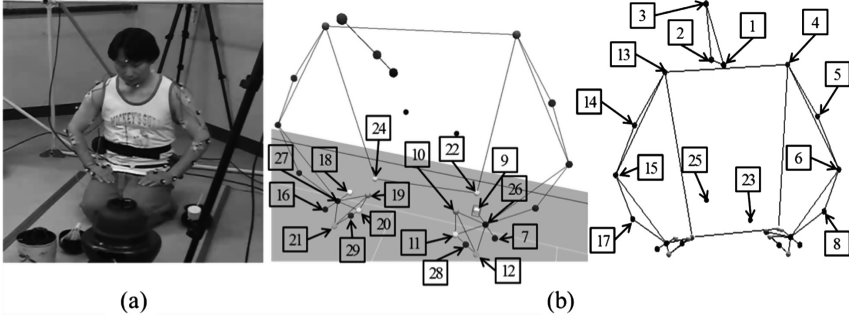


Fig. 1 Positions of markers

because the angles of body parts are considered to be more useful to represent fundamental movements and pauses. The radius $r_{i,n}$, inclination $\theta_{i,n}$, and azimuth $\varphi_{i,n}$ of the i th marker at the n th frame are given by Eqs. (2), (3), and (4), respectively. Note that the marker number i corresponds to the one shown in Fig. 1b.

$$r_{i,n} = \sqrt{x_{i,n}^2 + y_{i,n}^2 + z_{i,n}^2}. \quad (2)$$

$$\theta_{i,n} = \arccos\left(\frac{z_{i,n}}{r_{i,n}}\right). \quad (3)$$

$$\varphi_{i,n} = \arctan\left(\frac{y_{i,n}}{x_{i,n}}\right). \quad (4)$$

Based on this representation, the time series of the i th marker at the n th frame to the $(n - k + 1)$ th frame can be described as $(r_{i,n}, \theta_{i,n}, \varphi_{i,n}), \dots, (r_{i,n+k-1}, \theta_{i,n+k-1}, \varphi_{i,n+k-1})$, consisting of $3k$ values. This representation is used in our previous work [7].

The dimensionality reduction is performed by dividing the time series into S parts and quantizing each part. For example, when $k = 20$, $S = 4$, and $n = 1$, the time series is divided into four parts P_1 , P_2 , P_3 , and P_4 such that $P_1 = \{(r_{i,1}, \theta_{i,1}, \varphi_{i,1}), \dots, (r_{i,5}, \theta_{i,5}, \varphi_{i,5})\}$, $P_2 = \{(r_{i,6}, \theta_{i,6}, \varphi_{i,6}), \dots, (r_{i,10}, \theta_{i,10}, \varphi_{i,10})\}$, $P_3 = \{(r_{i,11}, \theta_{i,11}, \varphi_{i,11}), \dots, (r_{i,15}, \theta_{i,15}, \varphi_{i,15})\}$, and $P_4 = \{(r_{i,16}, \theta_{i,16}, \varphi_{i,16}), \dots, (r_{i,20}, \theta_{i,20}, \varphi_{i,20})\}$. We define feature values for the radius, inclination, and azimuth of each marker. A single feature value is represented as an $(S - 1)$ -digit ternary number (in this example, a 3-digit ternary number). That is, this time series data can be represented as 87 ternary numbers. This means that a fundamental movement or a pause is represented by an 87-dimensional feature vector.

The feature values are computed through a quantization process. The quantization is based on the values of the radius, inclination, and azimuth of each marker. A feature vector F is defined as Eq. (5).

$$F = (Q(r_1, \Theta_r), Q(\theta_1, \Theta_\theta), Q(\varphi_1, \Theta_\varphi), \dots, Q(r_{29}, \Theta_r), Q(\theta_{29}, \Theta_\theta), Q(\varphi_{29}, \Theta_\varphi)). \quad (5)$$

Here, the function Q is defined as Eq. (6).

$$Q(\alpha_i, \Theta) = \sum_{t=1}^{S-1} 3^t \cdot q(\mu(\alpha_i, t+1) - \mu(\alpha_i, t), \Theta). \quad (6)$$

In this equation, $\mu(\alpha_i, t)$ is the mean value of α_i of the frames included in P_i . Note that α_i can be the radius, inclination, or azimuth. That is, $\alpha_i \in \{r_i, \theta_i, \varphi_i\}$. For example, $\mu(r_i, 1)$ is the mean value of radius in P_1 , namely, $\{r_{1,1}, r_{1,2}, r_{1,3}, r_{1,4}, r_{1,5}\}$. The quantization function q is given by Eq. (7).

$$q(m, \Theta) = \begin{cases} 0 & \text{if } m \leq -\Theta \\ 1 & \text{else if } m \geq \Theta \\ 2 & \text{otherwise.} \end{cases} \quad (7)$$

The quantization function q returns 0, 1, or 2 depending on the value m and a threshold Θ . Note that Θ_r , Θ_θ , and Θ_φ are the thresholds for radius, inclination, and azimuth, respectively.

In the case that the radius continuously and fully increases, $\mu(r_i, t+1) - \mu(r_i, t)$ will be larger than the threshold Θ_r . The quantization function returns 1 in such a case. Conversely, when the radius continuously and fully decreases, the quantization function will return 0. If the change of radius is small, it will return 2. Therefore, the quantized value indicates the tendency of the change of radius. Of course, the tendency of the change of inclination and azimuth can also be obtained through the quantization function.

The function Q produces an $(S-1)$ -digit ternary number. For example, when $S = 4$, if the radius of the first marker constantly and fully decreases, q always returns 0 and thus Q yields a 3-digit ternary number 000_3 (equivalent to the decimal number 0). If the radius first increases, then decreases, and again increases, Q yields 101_3 (equivalent to the decimal number 10). The value Q returns ranges from 000_3 to 222_3 , namely, 0–26 in decimal representation. As a result, only 435 bits (approximately 55 bytes) are required to represent a fundamental movement or a pause since a 3-digit ternary number is described in 5 bits. This is a quite compact representation. Note that the size of a feature vector varies depending on the parameter of k and S . It is necessary to experimentally determine appropriate values of k and S .

In our previous work, we introduced similar quantization method [7]. However, its quantization function is different from that of the proposed method. The quantization function of the previous work uses only representative frames. A representative frame is the frame located in the center of each part P_i . As for the radius, for example, r_1 , r_2 , r_4 , and r_5 are not used to compute the feature value from P_1 since the representative frame in P_1 is the third frame. It seems to be sensitive to

outliers. We thus improved the feature value by using the mean value computed from all the frames in each part.

3.3 Feature Selection

The quantization-based dimensionality reduction considerably reduces the amount of motion data. However, the quantized data still have the redundancy because the movements (i.e., the time series) of some of markers are not so useful to describe fundamental movements and pauses. Removing the data of such markers leads to further conciseness. In addition, analyzing the usefulness of each marker will result in the discovery of important movement.

In this section, we propose a feature selection method by estimating the usefulness of the time series of the radius, inclination, and azimuth of each marker. Before the description of the feature selection method, let us explain the denotation.

- C ... Number of classes
The class means the type of fundamental movements. For example, when there are seven types of fundamental movements and a pause, the number of classes is eight ($C = 8$).
- T ... A training set
The usefulness of the time series is estimated through a kind of machine learning. A set of training examples is thus needed for the learning. Note that a training example is defined as a pair of a class label c ($c \in \{1, \dots, C\}$) and an 87-dimensional feature vector F described in Sect. 3.2.
- $N_{\alpha_i}(v, c)$... Number of training examples that the class labels are c and the feature values of α_i are v
Note that $v \in \{000_3, \dots, 222_3\}$ when $S = 4$. Since v ranges from 0 to $3^{S-1} - 1$ (in decimal representation), Eq. (8) holds for each α_i .

$$\sum_{c=1}^C \sum_{v=0}^{3^{S-1}-1} N_{\alpha_i}(v, c) = |T|. \quad (8)$$

The proposed method selects useful features on the basis of the distribution of N_{α_i} . N_{α_i} can be regarded as a two-dimensional histogram and $N_{\alpha_i}(v, c)$ corresponds to a bin of the histogram. Here, we define a normalized bin b_{α_i} by Eq. (9).

$$b_{\alpha_i}(v, c) = \frac{N_{\alpha_i}(v, c)}{\sum_{j=0}^{3^{S-1}-1} N_{\alpha_i}(j, c)}. \quad (9)$$

The denominator corresponds to the number of training examples having the class label c . This normalization mitigates the influence caused by the bias of the number of training examples among all the classes.

Based on this histogram, we define the usefulness U_{α_i} for α_i as Eq. (10).

$$U_{\alpha_i} = \frac{\sum_{v=0}^{3^{S-1}-1} B_{\alpha_i}(v)}{\sum_{v=0}^{3^{S-1}-1} I_{\alpha_i}(v)}. \quad (10)$$

In this equation, B_{α_i} and I_{α_i} are given by Eqs. (11) and (12), respectively.

$$B_{\alpha_i}(v) = \begin{cases} \sum_{j=1}^C \left(\max_c b_{\alpha_i}(v, c) - b_{\alpha_i}(v, j) \right) & \text{if } \max_c b_{\alpha_i}(v, c) > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

$$I_{\alpha_i}(v) = \begin{cases} 1 & \text{if } \max_c b_{\alpha_i}(v, c) > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

B_{α_i} indicates how useful α_i is. $B_{\alpha_i}(v)$ is maximized when the feature values of α_i of the training examples, which have a certain class label c , are v and those of the other training examples are other than v . In such a case, it is very easy to distinguish the examples having the class label c from the other examples. On the other hand, $B_{\alpha_i}(v)$ is minimized (becomes 0) when the numbers of the training examples, whose feature values of α_i are v , are same for all classes. In this case, it is impossible to accurately classify the examples.

If there is no example whose feature value of α_i is v , $B_{\alpha_i}(v)$ is defined as 0. Note that $\max_c b_{\alpha_i}(v, c)$ equals 0 in that case and that I_{α_i} is used to ignore that case.

We select the features depending on the values of the usefulness. In the case of selecting 10 features, for example, the features are sorted in descending order of the usefulness and then top 10 features are selected.

4 Experimental Evaluation

4.1 Overview of Experimental Evaluation

The experimental evaluation is conducted focusing on Japanese traditional tea ceremony. Several types of fundamental movements and pauses in Japanese traditional tea ceremony are described on the basis of the proposed representation. In order to evaluate the representational ability of the proposed method, we set up a classification problem and compute the classification accuracy.

The classification problem includes the discrimination of seven types of fundamental movements (shown in Fig. 2a–g) and the pauses (shown in Fig. 2h) that are considered to be important in Japanese traditional tea ceremony [4]. These

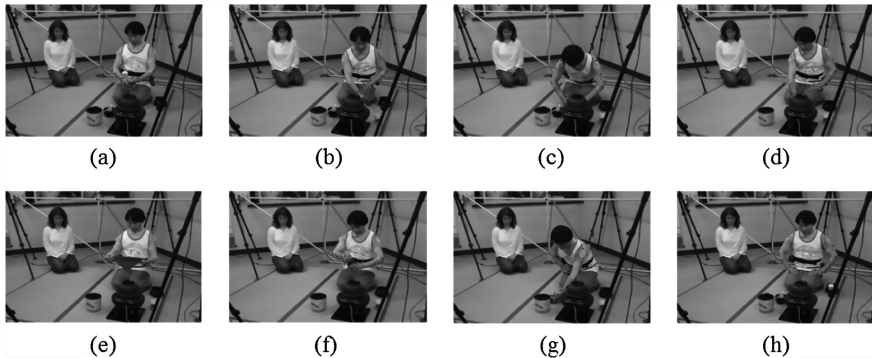


Fig. 2 Scenes in Japanese tea ceremony: **a** taking Hisyaku up, **b** putting Hisyaku down, **c** bowing, **d** putting a tea container down in front of a host, **e** Fukusa-sabaki, **f** purifying a tea container, **g** putting a tea container down in front of a water jug, and **h** an example of pause

movements and pauses are observed in the preparation phase of making a cup of tea. First, the host takes a Hisyaku, which is a ladle for scooping water, up (a), and puts it down (b). Next, he bows to the guest (c). He then puts a tea container down in front of him (d) and opens Fukusa, which is a sheet of cloth, to fold it (e). This action is called Fukusa-sabaki. He then purifies the tea container (f) and puts it down in front of a water jug (g). The pauses (h) appear several times typically between a movement and the subsequent movement.

4.2 Experimental Settings

The data set used in this experiment consists of 120 examples including eight classes as shown in Table 1. This is the same data set as used in our previous work [7].

The parameters k and S described in Sect. 3.2 were set to 40 and 4, respectively. The thresholds Θ_r , Θ_θ , and Θ_ϕ were set to 0.15, 0.0015, and 0.0015, respectively. These values were determined based on the result of the preliminary experiment, although the details of the preliminary experiment are omitted due to space limitation.

In order to construct a classification model, we used J48 implemented in Weka 3.6.13 [8]. It is the Weka implementation of C4.5 algorithm [9] which produces

Table 1 The numbers of examples of fundamental movements and pauses

Movement (Class)	Number of examples	Movement (Class)	Number of examples
(a)	13	(e)	37
(b)	24	(f)	10
(c)	7	(g)	9
(d)	11	(h)	9

decision trees. In this experiment, we restricted the type of decision trees to binary trees and performed a 10-fold cross validation.

4.3 Experimental Result

The classification result is evaluated by the F-measure of each fundamental movement (a)–(g) and the pause (h). The F-measure is computed by Eq. (13) using recall and precision defined as Eqs. (14) and (15), respectively. In these equations, X_c and x_c denote the set of examples whose class labels are c and that classified into the class c , respectively.

$$\text{F-measure} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}. \quad (13)$$

$$\text{recall} = \frac{|X_c \cap x_c|}{|X_c|}. \quad (14)$$

$$\text{precision} = \frac{|X_c \cap x_c|}{|x_c|}. \quad (15)$$

The average F-measure of the fundamental movements (a)–(g) and the pause (h) is shown in Fig. 3. The number of features selected varies from 1 to 87. Note that using 87 features corresponds to the case that the feature selection is not used, namely, all the features are used.

Figure 3 indicates that about half of features are redundant since the F-measure with 43 features (0.756) is very close to that with 87 features (0.768). This result shows the effectiveness of the feature selection.

Next, we compare the F-measure between the proposed method and our previous method [7]. The F-measure for each fundamental movement (a)–(g) and the pause

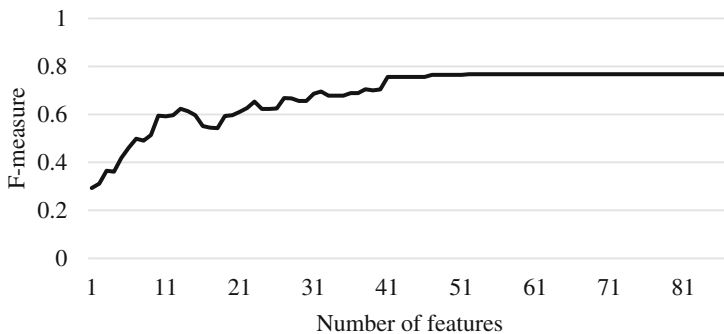


Fig. 3 Average F-measure

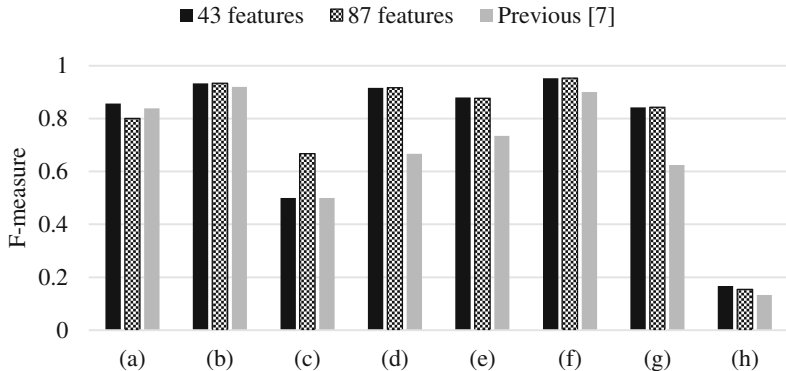


Fig. 4 F-measure for each fundamental movement and pause

(h) is shown in Fig. 4. In this figure, “43 features” and “87 features” means the F-measure of the proposed method with 43 selected features and with all 87 features, respectively. “Previous” means the F-measure of our previous work.

5 Consideration

5.1 Conciseness of Representation

As described in Sect. 3.2, the proposed method can represent a single fundamental movement or a pause in 435 bits without feature selection. This is fully concise but the required data size is the same as that of our previous method [7]. For further conciseness, we propose the feature selection method.

The experimental result shown in Fig. 3 shows that using half of features is sufficient. Therefore, the data size can be reduced to 215 bits. This can be a remarkable advantage for developing large-scale archives.

5.2 Representational Ability

Figure 4 indicates that the proposed method outperforms our previous method in all the fundamental movements and the pause. This is due to the improvement of the quantization method. However, the F-measure of the pause is still very low. This problem should be resolved in the future work.

Comparing the result with and without feature selection, the F-measure does not decrease except for the fundamental movement (c). This fundamental movement is different from the others in the point that no instrument (such as Hisyaku and

Fukusa) is used in the movement. Some of the features eliminated by the feature selection may be needed for this type of fundamental movement.

5.3 Computational Efficiency

We make a comparison between the proposed method and our previous method [7] in the time required to perform the experiment described in Sect. 4.

When the feature selection is not used, the time required by the proposed method is approximately 103 % of the time required by our previous method using a commonly-used personal computer. The required time slightly increases because of the difference of the quantization method.

By introducing the feature selection, the time required by the proposed method (using 43 selected features) is approximately 97 % of the time required by our previous method. The reduction of the number of features leads to the decrease of required time. The feature selection is, therefore, effective to improve computational efficiency as well.

5.4 Useful Features

The feature selection not only improves the computational efficiency but also clarifies the usefulness of each feature. Table 2 shows the top 10 features of the usefulness. All of them are related to the angle. Specifically, the azimuth tends to be regarded as useful features. As for the positions of the markers, six features out of ten features are associated with left arm or left hand. Considering this result, analyzing angular movement around left arm or left hand may lead to the understanding of the important movements in Japanese traditional tea ceremony.

Table 2 Usefulness of top 10 features

Rank	Feature	Usefulness	Rank	Feature	Usefulness
1	θ_8	2.191	6	θ_9	1.928
2	φ_{28}	2.085	7	θ_7	1.870
3	φ_{23}	1.988	8	φ_3	1.808
4	φ_4	1.987	9	φ_{25}	1.679
5	φ_{29}	1.951	10	φ_6	1.641

6 Conclusion

For the purpose of archiving traditional skills, an efficient representation of the fundamental movements and pauses in the traditional skills was proposed. The proposed representation is quite compact due to the dimensionality reduction of the motion data. In addition, we proposed a feature selection method to further reduce the amount of data and to discover important features.

The experimental result using the data set of Japanese traditional tea ceremony clarified that the representational ability of the proposed method is better than that of our previous work. Moreover, we confirmed that the feature selection made the computation more efficient compared with the previous work. The result of the feature selection indicated that the angular movements around left arm and left hand were discriminative. This could be a clue to the understanding of the important movements in Japanese traditional tea ceremony.

Although most of the fundamental movement were appropriately represented by the proposed method, the pause was not precisely represented. Improving the representation of the pause is included in the future work. Additionally, evaluating the proposed method for a wide variety of traditional skills is also in the future work.

Acknowledgments We thank Dr. Masashi Kume (Kyoto Bunkyo Junior College) and Dr. Tetsuya Yoshida (Kyoto Institute of Technology) for permitting us to use the experimental motion data, where Mr. Toru Ota (Kyoto Institute of Technology) acted as the host and Ms. Kristin Surak (University of California, Los Angeles) acted as the guest. We also give great thanks to them.

This research is partially supported by Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B), 15H02769, 2015-2017.

References

1. Morimoto, K., Kuwahara, N.: An analysis of body movements to sharpen cutlery by Kyoto Bow Artisan. In: Proceedings of International Conference on Advanced Applied Informatics, pp. 385–388 (2013)
2. Kume, M., Yoshida, T.: Characteristics of technique of skill in traditional craft workers in Japan. In: Proceedings of International Conference on Affective and Pleasurable Design, pp. 5289–5297 (2012)
3. Morimoto, K., Kuwahara, N.: Holistic analysis on affective source of Japanese traditional skills. In: Proceedings of International Conference on Affective and Pleasurable Design, pp. 5239–5243 (2012)
4. Hochin, T., Nomiya, H.: Deriving pauses for obtaining fundamental movements in traditional skills. In: Lee, R.Y. (ed.) Applied Computing and Information Technology: Studies in Computational Intelligence, vol. 533, pp. 19–30. Springer, Berlin (2014)
5. Kuramoto, I., Nishimura, Y., Yamamoto, K., Shibuya, Y., Tsujino, Y.: Visualizing velocity and acceleration on augmented practice mirror self-learning support system of physical motion. In: Proceedings of International Conference on Advanced Applied Informatics, pp. 365–368 (2013)
6. Kim, S.-W., Park, S., Chu, W.W.: An index-based approach for similarity search supporting time warping in large sequence databases. In: Proceedings of International Conference on Data Engineering, pp. 607–614 (2005)

7. Hochin, T., Nomiya, H.: Considerations on archiving traditional skills focusing on pauses and fundamental movements. In: Proceedings of 6th International Conference on Applied Human Factors and Ergonomics, pp. 4566–4573 (2015)
8. Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>
9. Quinlan, J. R.: C4.5: Programs for Machine Learning. Morgan Kaufman, San Francisco (1993)

Advances in Affective and Pleasurable Design
Proceedings of the AHFE 2016 International
Conference on Affective and Pleasurable Design, July
27-31, 2016, Walt Disney World®, Florida, USA
Chung, W.; Shin, C.S. (Eds.)
2017, XV, 714 p. 329 illus., 252 illus. in color., Softcover
ISBN: 978-3-319-41660-1