Paper:

Acquisition of Embodied Knowledge on Gesture Motion by Singular Value Decomposition

Isao Hayashi*, **Yinlai Jiang****, **and Shuoyu Wang****

*Graduate School of Informatics, Kansai University
2-1-1 Ryozenji-cho, Takatsuki, Osaka 569-1095, Japan
E-mail: ihaya@cbii.kutc.kansai-u.ac.jp
**School of Systems Engineering, Kochi University of Technology
185 Miyanokuti, Tosayamada, Kami, Kochi 782-8502, Japan
[Received March 10, 2011; accepted July 15, 2011]

Communication is classified in terms of verbal and nonverbal information. We discuss an acquisition method of knowledge from nonverbal information. In particular, a gesture is an efficient form of nonverbal communication as well as in verbal ways, and we formulate here a method that measures similarity and estimation between gestures. A gesture includes human embodied knowledge, and therefore the visible bodily actions can communicate particular messages. However, we have infinite patterns for gesture, determined by personality. Recently, the singular spectrum analysis method is utilized as an attractive method. In this paper, we propose a new method for acquiring embodied knowledge from time-series data on gestures using singular value decomposition. The motion behavior is categorized into several clusters with similarity and estimation between interval time-series data. We discuss the usefulness of the proposed method using an example of gesture motion.

Keywords: human embodied knowledge, skill acquisition, motion recognition, singular value decomposition

1. Introduction

Human-friendly robots are expected to support future daily life at home, office, and medical treatment, and assure safety and security in natural disasters or large-scale accident [1]. Unlike industrial robots working in factories, human-friendly robots must be able to communicate in ways similar to how persons communicates, and such how to communicate can allow us to avoid fear for robot. The communication between human and robot is classified into verbal and nonverbal information. Persons communicate in nonverbal ways such as gestures, facial expression, posture, appearance, listening, and eye contact, as well as in verbal ways, such as speaking and writing. We have been studying models to acquire knowledge from nonverbal information of robot. In particular, a gesture is an efficient form of nonverbal communication, and so we are focusing clearly on methods that acquire knowledge from human gestures for teaching them to robots.

To develop human-friendly robots, we need inherent skill ability to internalize in a body as embodied knowledge by external action [2–9], e.g., how to acquire skill while moving economically as athletes, how to make movement more flexible as martial artists, how to uplift feeling as musicians, and how to acquire specific finger sensitivity as ceramists, or how to move attractively as dancers. Such skill is a native human endowment and we should implement this skill as knowledge embodied in a robot.

In this paper, we propose a new method to acquire embodied knowledge of gesture movement [10, 11]. Gesture recognition has been studied extensively, and there have been varied approaches to handling gesture recognition. Mitra et al. [12] surveyed methods of gesture recognition, and applied the Hidden Markov model to gesture motion data. Iwai et al. [13] proposed a CAD method to recognize hand gestures from a monocular image detected using colored gloves. Lamar et al. [14] proposed a neural network, Temporal-CombNET (T-CombNET), which deals with time series and classification, and applied it to Japanese-Kana finger spelling recognition. Jerde et al. [15] used an instrumented glove to measure the angular positions of hand joints for recognizing fingerspelling hand shapes, and reduced hand dimensionality by discriminant analysis using Principal Component Analysis (PCA). Suk et al. [16] proposed gesture recognition model for recognizing hand gestures in video streaming using dynamic Bayesian network framework or DBN model. In addition, several other methods has been proposed [17, 18]. The Hidden Markov model is not, however, effective if the number of states is large, e.g., that data includes a different long period and discontinuous data. Since neural network is too sensitive in time-series data length, the accuracy is not so well. Principal Component Analysis reduces the number of explanatory variables, and is a kind of model for visualization with principal component variables. It is possible however to lose significant principal component variables when the proportion of variance is low and the number of the data is inadequate.

In our approach [19], we discuss a new gesture recognition method to identify 3-dimensional gesture motions using Singular Value Decomposition (SVD), in which left and right singular vectors, and singular values are decomposed from the Hankel matrix. Since the left singular vector represents Hankel matrix characteristics and the singular value means the strength of the left singular vector, it is used more generally as a method for extracting characteristics from observed data. When the Hankel matrix is defined from time-series data, the method to find dominant components using singular value decomposition on Hankel matrices is called Singular Spectrum Analysis. Recently, the singular spectrum analysis is used in time-series data analysis for data mining [20] and motion analysis to extract similarities and differences in human behavior [21]. Applications using Singular Value Decomposition include computing the pseudo inverse, least squares data fitting, matrix approximation, and determining the rank, range, and null space of a matrix [22].

In our proposed model, we formulate the criteria of similarity and estimation between gestures and we distinguish among gestures in dimension space of left singular vectors. To formulate criteria, we proposed two kinds of methods, first method to measure the similarity between gesture distances and the second method to measure the similarity of the gesture vector. We applied the proposed methods to five gestures, and discuss the usefulness of the proposed methods by the motion examples.

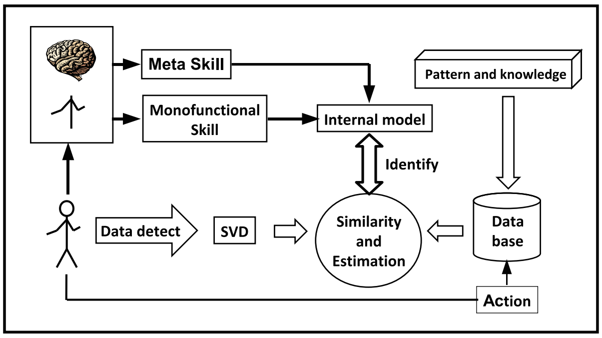

Fig. 1. Proposed model.

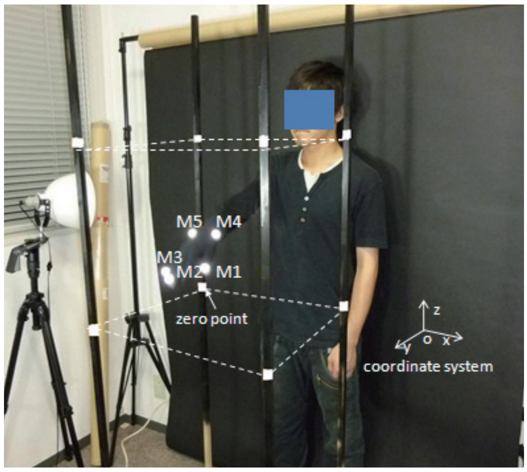

Fig. 2. Experiment environment.

2. 3-Dimensional Gesture Measurement

Our approach assumes that skill ability consists of hierarchical structure with a mono-functional layer to generate the single function result and a meta-functional layer that adapted itself to environmental change as shown in **Fig. 1**. We assume that skill is knowledge embodied in the human body. Therefore, we attach an action tag to every gesture after having detected movement and identify relations between gesture data and action tags as an internal model using Singular Value Decomposition. In general, supervisor signal does not exist to the physical movement, but it is said that we acquire a calculation model in brain instead of supervisor signal as to environmental input and output which surround the brain [23]. This computational model is called an internal model, for which Kawato [24] has argued that inverse model with feedback and feedforward control is useful in modeling motor control. Miall et al. [25] proposed a forward model to output a signal to the controlled object via feedback control using the Smith predictor having no time delay for the control. The forward and inverse models are remarkable in constituting an internal model. Instead of arguing about internal model construction, we propose identifying the relationship between sensor input and the output of internal model movement as knowledge. We propose two methods to identify of the internal model, measuring the similarity between gesture distances and measuring the similarity between gesture vectors. Since these two identification models are data-dependent, the priority of identification model use depends on the data.

Hand gestures are measured with Movetr/3D and GE60/W (Library, Tokyo, Japan). Subjects are two men, SW and ST, in their 20s. Five markers were measured, M_1 at the tip of the thumb, M_2 on the tip of the middle finger, M_3 at the tip of the little finger, M_4 at the thumb-side of the wrist, and M_5 at the little finger side of the wrist. Gestures were performed in 50 cm $\times$ 50 cm $\times$ 50 cm cubic space, whose zero point and coordinates are shown in **Fig. 2**. In experiments, five hand gestures – Come Here (CH), Go Away (GA), Go Right (GR), Go Left (GL) and Calm Down (CD) – were performed by the 2 subjects, and 9 times each by each subject. Data on the first 5 times was used to acquire gesture pattern. Data on last 4 times was used to be distinguished.

The measurement time-series data of M_2 when subject SW performed the five gestures is shown in **Fig. 3**. Five hand gestures are shown in **Fig. 4**. In **Fig. 3**, movement change for GA, CH and CD is big at top and bottom onto the z-axis and in front and back onto the y-axis. For GR and GL, movement change is big left and right onto the x-axis.

3. Gesture Analysis Using Singular Value Decomposition

Suppose that the hand has w measurement points $(P_1, P_2, \ldots, P_w)$. At point P_i, the measured data series of

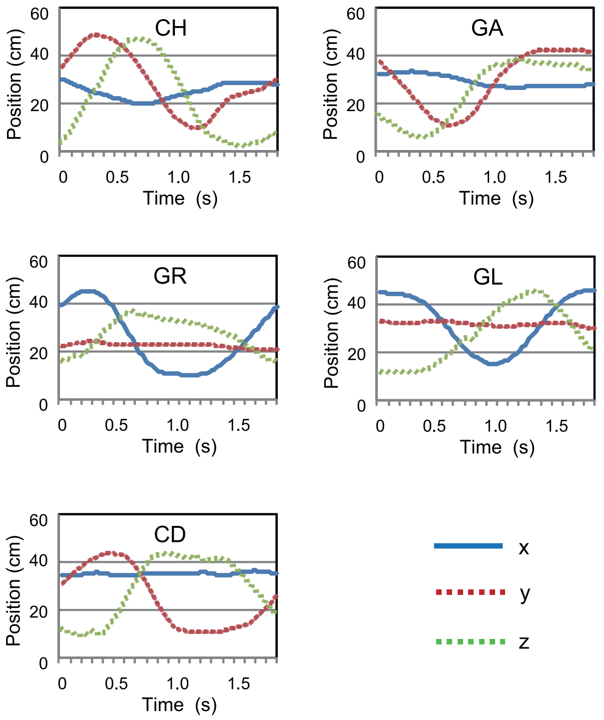

Fig. 3. Time-series data of gestures.

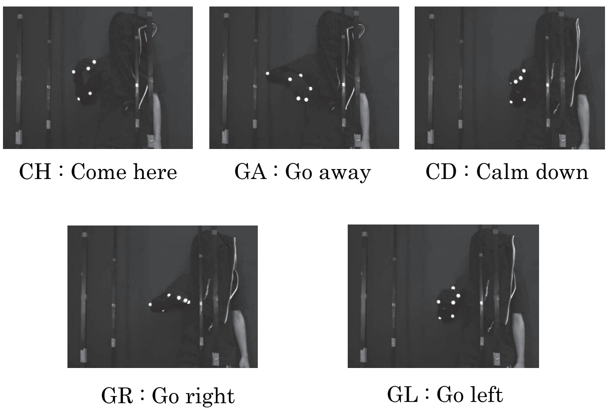

Fig. 4. Five kinds of hand gestures.

gesture G is denoted as $\tau^{i,G}$. Data series $\tau^{i,G}$ consists of 3-dimensional data $(X^{i,G}, Y^{i,G}, Z^{i,G})$. Time-series data $\tau^{i,G} = (X^{i,G}, Y^{i,G}, Z^{i,G})$ is decomposed into n vectors by m data sampling by overlapping, and matrices $M_X^{i,G}, M_Y^{i,G}$ and $M_Z^{i,G}$ are constructed as a collective of the change in X, Y and Z coordinates of gestures. The design of how to construct matrix $M_X^{i,G}$ is shown in **Fig. 5**. Matrices $M_X^{i,G}, M_Y^{i,G}$, and $M_Z^{i,G}$ are expressed as follows:

$$M_X^{i,G} = (X_1^{i,G}, X_2^{i,G}, \ldots, X_n^{i,G})^T \quad \ldots\ldots\ (1)$$

$$M_Y^{i,G} = (Y_1^{i,G}, Y_2^{i,G}, \ldots, Y_n^{i,G})^T \quad \ldots\ldots\ (2)$$

$$M_Z^{i,G} = (Z_1^{i,G}, Z_2^{i,G}, \ldots, Z_n^{i,G})^T \quad \ldots\ldots\ (3)$$

where $X_p^{i,G} = (x_{p,1}^{i,G}, x_{p,2}^{i,G}, \ldots, x_{p,m}^{i,G})$, $p = 1, 2, \ldots, n$, and x is data on X coordinates. We define $Y_p^{i,G}$ and $Z_p^{i,G}$ the same way.

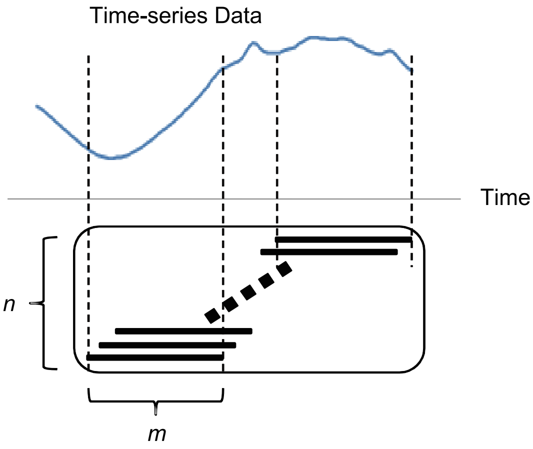

Fig. 5. Design of matrix $M_X^{i,G}$.

Suppose $M_k^{i,G}$, $k = \{X, Y, Z\}$ is an m-by-n matrix as the general format of $M_X^{i,G}, M_Y^{i,G}, M_Z^{i,G}$. The factorization of the form is as follows:

$$M_k^{i,G} = U_k^{i,G} \Sigma_k^{i,G} \{V_k^{i,G}\}^T \quad \ldots\ldots\ldots\ (4)$$

where $U_k^{i,G} = (u_{1,k}^{i,G}, u_{2,k}^{i,G}, \ldots, u_{m,k}^{i,G})$ is an m-by-m unitary matrix, $\{V_k^{i,G}\}^T$ the conjugate transposition of $V_k^{i,G} = (v_{1,k}^{i,G}, v_{2,k}^{i,G}, \ldots, v_{n,k}^{i,G})$, and an n-by-n unitary matrix, and matrix $\Sigma_k^{i,G}$ a m-by-n diagonal matrix with nonnegative real numbers on the diagonal. Such factorization is called Singular Value Decomposition of $M_k^{i,G}$. Diagonal entries of $\Sigma_k^{i,G}$ are known as singular values of $M_k^{i,G}$. Matrix $U_k^{i,G}$ contains the left singular vectors of $M_k^{i,G}$ and matrix $V_k^{i,G}$ contains the right singular vectors of $M_k^{i,G}$.

Now, we choose $M_X^{i,G}$ as an example of matrix $M_k^{i,G}$ and discuss gesture analysis. Matrix $M_X^{i,G}$ is decomposed into the product of $U_X^{i,G}$, $\Sigma_X^{i,G}$, and $V_X^{i,G}$. Intuitively, left singular vectors in $U_X^{i,G}$ form a set of patterns of $M_X^{i,G}$ and diagonal values in matrix $\Sigma_X^{i,G}$ are singular values, thought of as scalar, by which each corresponding left singular vector affects matrix $M_X^{i,G}$. Let us denote the number of left singular vectors as l, and the element number of the j-th left singular vector as q. We define pairs of singular values and the left singular vector as $((\sigma_{1,X}^{i,G}, u_{1,X}^{i,G}), (\sigma_{2,X}^{i,G}, u_{2,X}^{i,G}), \ldots, (\sigma_{l,X}^{i,G}, u_{l,X}^{i,G}))$, for $u_{j,X}^{i,G} = (\hat{u}_{1j,X}^{i,G}, \hat{u}_{2j,X}^{i,G}, \ldots, \hat{u}_{hj,X}^{i,G}, \ldots, \hat{u}_{qj,X}^{i,G})$ in descending order of singular values, where $\hat{u}_{hj,X}^{i,G}$ is the h-th element of the j-th left singular vector $u_{j,X}^{i,G}$. The left singular vector corresponding to the singular value expresses the whole time-series data characteristic of better if a singular value is larger. Therefore, left singular vectors $(u_{1,X}^{i,G}, u_{2,X}^{i,G}, \ldots, u_{l,X}^{i,G})$ represent hand gesture characteristic of well.

3.1. Similarity Between Gesture Distances

Suppose that observed data series are divided into $\tau_{TRD}^{i,G}$ for training data series, and $\tau_{CHD}^{i,G}$ for checking data. Let us denote left singular vector $\tau_{X,TRD}^{i,G}$ for training data related to X coordinate values of point P_i on the hand for gesture G as $U_{X,TRD}^{i,G} = (u_{1,X,TRD}^{i,G}, u_{2,X,TRD}^{i,G}, \ldots, u_{l,X,TRD}^{i,G})$ for $u_{j,X,TRD}^{i,G} = (\hat{u}_{1j,X,TRD}^{i,G}, \hat{u}_{2j,X,TRD}^{i,G}, \ldots, \hat{u}_{hj,X,TRD}^{i,G}, \ldots, \hat{u}_{qj,X,TRD}^{i,G})$.

We define left singular vectors $\tau_{X,CHD}^{i}$ for checking data as $U_{X,CHD}^{i} = (u_{1,X,CHD}^{i}, u_{2,X,CHD}^{i}, \ldots, u_{l,X,CHD}^{i})$, for $u_{j,X,CHD}^{i} = (\hat{u}_{1j,X,CHD}^{i}, \hat{u}_{2j,X,CHD}^{i}, \ldots, \hat{u}_{hj,X,CHD}^{i}, \ldots, \hat{u}_{qj,X,CHD}^{i})$ in the same way.

To recognize hand gestures, three similarity criteria are defined in data $\tau^{i,G} = (X^{i,G}, Y^{i,G}, Z^{i,G})$ as follows:

$$S_1 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sum_{j=1}^{l}\left|\sum_{h=1}^{q}\hat{u}_{hj,k,TRD}^{i,G} - \sum_{h=1}^{q}\hat{u}_{hj,k,CHD}^{i}\right| \quad . \quad (5)$$

$$S_2 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sum_{j=1}^{l}\sum_{h=1}^{q}\left|\hat{u}_{hj,k,TRD}^{i,G} - \hat{u}_{hj,k,CHD}^{i}\right| \quad . \quad . \quad (6)$$

$$S_3 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sqrt{\sum_{j=1}^{l}\sum_{h=1}^{q}(\hat{u}_{hj,k,TRD}^{i,G} - \hat{u}_{hj,k,CHD}^{i})^2} \quad . \quad (7)$$

Similarity S_1 is defined by the absolute differential of the total left singular vector between training and checking data. Similarity S_2 is defined by the total absolute differential of left singular vectors between training and checking data in the same order. Similarity S_3 is defined by the Euclidean distance between left singular vectors of training and checking data in multidimensional space.

Since there are w measurement points $(P_1, P_2, \ldots, P_w)$, estimated gesture G^* is identified by the following two estimations:

$$E_1 : G^* = \{G_f | \max_f \sum_{i=1}^{w} n(G_f^i),$$
$$for\ G_f^i = \{G_f | \min_f r_i(U_{TRD}^{i,G_f}, U_{CHD}^{i})\}\} \quad . \quad (8)$$

$$E_2 : G^* = \{G_f | \min_f \sum_{i=1}^{w} r_i(U_{TRD}^{i,G_f}, U_{CHD}^{i})\} \quad . \quad . \quad (9)$$

where G_f is the f-th gesture among five hand gestures and $n(G_f^i)$ is a counting function that is $n(G_f^i) = 1$ if condition G_f^i is satisfied at point P_i on the hand.

Estimation E_1 is defined for a gesture by counting the large number of minimized similarity values. Estimation E_2 is defined by minimized total similarity values.

3.2. Similarity Between Gesture Vector

As for the second method, gesture recognition is based on the similarity between gesture vectors. Now, consider the data series in $M_X^{i,G}$ of Eq. (1). If one part of the data series $X_p^{i,G}$ is replaced by another data series X_{CHD}^{i}, singular values and left singular vectors of $M_{X,CHD}^{i}$, differ from those of $M_X^{i,G}$.

$$M_{X,CHD}^{i} = (X_1^{i,G}, X_2^{i,G}, \ldots, X_{p-1}^{i,G}, X_{CHD}^{i}, X_{p+1}^{i,G}, \ldots, X_n^{i,G})^T. \quad (10)$$

The difference between left singular vectors of X, $M_X^{i,G}$ and $M_{X,CHD}^{i}$, is determined by how X_{CHD}^{i} differs from other $p-1$ data series. $M_{X,CHD}^{i}$ patterns change more when $X_p^{i,G}$ is replaced by many dissimilar data series than when it is replaced by a similar one, so the difference between left singular vectors of $M_X^{i,G}$ and X_{CHD}^{i} are considered a measure of the difference between X_{CHD}^{i} and other data series. If X_{CHD}^{i} comes from another hand gesture, the difference is used as a criterion for judging whether X_{CHD}^{i} comes from the same hand gesture as other data series. For our algorithm, the location of X_{CHD}^{i} is fixed at the data series end. Therefore, the n-th $X_n^{i,G}$ is only replaced by another data series X_{CHD}^{i} in Eq. (10).

To recognize hand gestures, three similarity criteria are defined in data $\tau^{i,G} = (X^{i,G}, Y^{i,G}, Z^{i,G})$ as follows:

$$S_4 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sum_{j=1}^{l}\left|\sum_{h=1}^{q}\hat{u}_{hj,k,TRD}^{i,G} - \sum_{h=1}^{q}\hat{u}_{hj,k,CHD}^{i}\right| \quad (11)$$

$$S_5 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sum_{j=1}^{l}\sum_{h=1}^{q}\left|\hat{u}_{hj,k,TRD}^{i,G} - \hat{u}_{hj,k,CHD}^{i}\right| \quad . \quad (12)$$

$$S_6 : r_i(U_{TRD}^{i,G}, U_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sqrt{\sum_{j=1}^{l}\sum_{h=1}^{q}(\hat{u}_{hj,k,TRD}^{i,G} - \hat{u}_{hj,k,CHD}^{i})^2} \quad (13)$$

Since there are w measurement points $(P_1, P_2, \ldots, P_w)$, estimated gesture G^* is identified by the following estimation:

$$E_3 : G^* = \{G_f | \min_f \sum_{i=1}^{w} r_i(u_{TRD}^{i,G_f}, u_{CHD}^{i})\} \quad . \quad . \quad (14)$$

In our proposed method, all time series must have the same number of data to compose matrix $M_X^{i,G}$ as shown in Eq. (1). The length of hand gestures, however, in real measurement differs with the gestures and the subject that performs them. Therefore, preprocessing is necessary to make the data series have the same number of data. In this paper, the number of data was set to be the average number. If a data series contains more data than the average number, data is deleted from the data series at the same interval. If a data series contains fewer data than the average number, data is interpolated into the data series at the same interval. Interpolated data is calculated using quadric interpolation.

Table 1. Recognition of gestures.

	Similarity (S_1)	Similarity (S_2)	Similarity (S_3)
Estimation (E_1)	70.0	90.0	80.0
Estimation (E_2)	60.0	80.0	80.0

Table 2. Result of subject SW gestures.

Estimated Gesture	Gesture of SW				
	CH	GA	GR	GL	CD
CH	12	1	4	4	4
GA	0	10	2	2	4
GR	0	0	0	4	0
GL	2	1	9	5	0
CD	1	3	0	0	7
Result	CH	GA	GL	GL	CD

Table 3. Result of subject ST gestures.

Estimated Gesture	Gesture of ST				
	CH	GA	GR	GL	CD
CH	8	2	3	2	0
GA	1	9	1	0	2
GR	0	0	11	1	0
GL	3	2	0	10	2
CD	3	2	0	2	11
Result	CH	GA	GR	GL	CD

Table 4. Comparison of markers by gesture distances.

Sub./Ges.	Markers				
TW	M_1	M_2	M_3	M_4	M_5
CH	4/GA	4/GA	4/CD	4/GA	4/GA
GA	8/GA	9/GA	9/GA	8/GA	8/GA
GR	7/GR	8/GR	8/GR	6/GR	5/GR*
GL	5/GL	4/GL*	4/GL*	5/GL	5/GL
CD	6/CD	5/GA	5/GA	7/CD	8/CD
ST	M_1	M_2	M_3	M_4	M_5
CH	8/CH	9/CH	11/CH	9/CH	7/CH
GA	7/GA	9/GA	9/GA	9/GA	5/CD
GR	9/GR	10/GR	11/GR	6/GR	9/GR
GL	7/GL	8/GL	10/GL	5/GL	8/GL
CD	4/CD*	5/GL	5/GL	5/CD	6/CD
Accuracy (%)	93.85	80.28	81.58	93.75	86.15

4. Results

In order to show the usefulness of the proposed methods, we distinguished between gestures of two subjects SW and ST using two methods. Subjects only numbered two and the number for experiments was too few, but we need to conduct experiments with a few people too much to reduce the difference between subjects. We requested each subject to repeat movements nine times and discussed the usefulness of the proposed methods based on observed data. The average number of SW gesture data was 100.3 (SD: 12.8) and that of ST gestures was 150.3 (SD: 16.4). The average number of SW and ST gestures was 125.2 (SD: 29.0). The number of data m was set to be 125. We set $n = 5$, $q = 125$, $l = 1$ and $w = 5$.

4.1. Similarity Between Gesture Distances

Table 1 shows recognition results for gesture patterns based on the three different similarities – Eqs. (5), (6) and (7). Recognition results suggest that similarity S_1 and S_2 leaded to relatively higher correct recognition rates. The pair of similarity S_2 and estimation E_1 is 90.0%, and results suggest that the pair S_2 and E_1 is more feasible in gesture recognition. **Tables 2** and **3** show the counting the number of measurement points for pair S_2 and E_1. For example, 15 executions of CH of SW were recognized as CH 12, GL twice and CD once, based on pair S_2 and E_1 in **Table 2**. Recognition results suggest that gestures CH, GA and CD are distinguished between well, but not so well for GR and GL of SW because ST tried to make gestures more emphatic and melodramatic than SW. In general, it is difficult to distinguish between gestures GR and GL, so these results are understandable.

Next, by calculating left singular vectors at each marker, we compared the accuracy of markers for gesture recognition with similarity S_2 and estimation E_1. **Table 4** shows the large number of minimized similarity values. **Table 4** shows estimated gestures and the number of times recognized for each marker and every gesture in a/b, where b expresses a recognized gesture, and a expresses the number of times. In addition, b^* shows that other estimated gestures exist as well as b. As a result, the first M_1 was selected as the most important marker because the accuracy is highest, at 93.85%. Since first marker M_1 measures time series at the tip of the thumb and it is largely related to thumb movement for gesture recognition, the result is understandable.

4.2. Similarity Between Gesture Vectors

Table 5 shows recognition results based on the three similarity definitions S_4, S_5, and S_6, in Eqs. (11), (12) and (13). Recognition results suggest that similarity definitions of S_5 and S_6 led to relatively higher correct recognition rates while the correct rate of recognition based on S_4 was very low, making S_5 and S_6 more feasible in gesture recognition.

To compare markers by similarity between gesture distances, we compared the accuracy of markers by similarity between gesture vectors with similarity S_5 and estimation E_3. **Table 6** shows the large number of minimized similarity values. The first M_1 was selected as the most important marker, the same as for **Table 4**. We realize that thumb movement is strongly related to gesture recognition by these result.

Table 5. Results of method for similarity between gesture vectors.

Sub.	Ges.	Similarity (S_4)	Similarity (S_5)	Similarity (S_6)
TW	CH	1	2	1
	GA	3	4	4
	GR	2	3	3
	GL	1	3	3
	CD	0	3	3
ST	CH	2	2	2
	GA	1	4	4
	GR	3	4	4
	GL	0	3	2
	CD	0	4	4
Accuracy (%)		30.3	80.0	75.0

Table 6. Comparison of markers by gesture vectors.

Sub./Ges.	Markers				
TW	M_1	M_2	M_3	M_4	M_5
CH	2/GA	2/GA	2/GA	3/GA	2/GA
GA	4/GA	4/GA	2/GA*	4/GA	4/GA
GR	3/GR	3/GR	3/GR	3/GR	3/GR*
GL	3/GL	3/GL	3/GL	3/GL	3/GL
CD	3/CD	2/GA	3/GA	2/CD	2/CD
ST	M_1	M_2	M_3	M_4	M_5
CH	2/CH*	2/CH*	2/CH*	2/CH*	2/CH*
GA	4/GA	4/GA	4/GA	4/GA	4/CD
GR	4/GR	4/GR	4/GR	3/GR	3/GL
GL	4/GL	4/GL	4/GL	3/GR	4/GR
CD	4/CD	4/CD	4/CD	3/CD	4/CD
Accuracy (%)	93.94	87.50	83.87	80.00	58.06

Finally, we compared our proposed method and the conventional method that calculates correlation coefficient and Principal Component Analysis (PCA) based on gesture time series. **Table 7** shows results for five methods. We used the gesture distance that distinguishes gesture movement with similarity S_2 and estimation E_1, and the gesture vector that distinguishes gesture movement with similarity S_5 and estimation E_3. For correlation coefficients, the following combination of similarity S_7 and estimation E_4 was used to calculate recognition accuracy. The following combination of similarity S_8 and estimation E_5 was used for clustering by principal component analysis, and the combination of similarity S_9 and estimation E_5 is used for distance method by principal component analysis:

$$S_7 : r_i(U^{i,G}_{TRD}, U^i_{CHD}) = \frac{1}{3}\sum_{k=1}^{3} R(d^{i,G}_{TRD,k}, d^i_{CHD,k}) \quad \ldots\ldots \quad (15)$$

$$S_8 : r_i(U^{i,G}_{TRD}, U^i_{CHD}) = \min_f \sqrt{\sum_{c=1}^{H} (f_{XYZ\to c}(E(d^{i,G_f}_{TRD})) - f_{XYZ\to c}(E(d^i_{CHD})))^2} \quad \ldots\ldots \quad (16)$$

$$S_9 : r_i(U^{i,G}_{TRD}, U^i_{CHD}) = \min_f \sqrt{\sum_{c=1}^{H} (f_{G_f\to c}(d^{i,G_f}_{TRD}) - f_{G_f\to c}(d^i_{CHD}))^2} \quad \ldots \quad (17)$$

$$E_4 : G^* = \{G_f | \max_f \frac{1}{w}\sum_{i=1}^{w} r_i(U^{i,G_f}_{TRD}, U^i_{CHD})\} \quad (18)$$

$$E_5 : G^* = \{G_f | \max_f \sum_{i=1}^{w} n(r_i(U^{i,G_f}_{TRD}, U^i_{CHD}))\} \quad (19)$$

where $d^{i,G}_{TRD,k}$, and d^i_{CHD} are time series of training and checking data and $R(x,y)$ means the correlation coefficient between x and y. In addition, $E(\ )$ means an average, $f_{A\to B}(\ )$ a function of principal component analysis that converts from variable A to principal axis B, and H the number of principal axes where we set $H = 3$.

Table 7. Comparison of methods.

Markers	Ges. Dist.	Ges. Vec.	Corre. Co.	PCA Clus.	PCA Dist.
M_1	93.85	93.94	60.0	20.0	60.0
M_2	80.28	87.5	80.0	40.0	80.0
M_3	81.58	83.87	80.0	60.0	40.0
M_4	93.75	80.0	80.0	40.0	40.0
M_5	86.15	58.06	60.0	20.0	40.0
Average	87.12	80.67	72.0	36.0	52.0
All Markers	90.0	80.0	70.0	60.0	40.0

Since the gesture distance method ensures 90.0% of recognition accuracy when all markers are used, and 87.12% in the average recognition of each marker as shown in **Table 7**, this means that gesture distance is a useful model in motion recognition. Since the recognition of the first marker of the gesture vector is 93.94%, and compared to 80.0% for all markers from **Table 7**, the gesture vector is considered a useful model. Recognition accuracy using the correlation coefficient is not worst, but we cannot say that the model is good. Furthermore, the cluster model using the principal component analysis (PCA Clustering) and the model comparing distances between principal axes (PCA Distance) did not provide high accuracy. We consider this to be due to the two data characteristics of the different length of time-series data and the similar trajectory between gesture movements. For Singular Value Decomposition, we consider that both methods gesture distance and the gesture vector ensure high recognition accuracy by extracting a movement characteristic regardless of time-series data length.

5. Discussion

Similar to speech and handwriting, gestures vary between individuals, but do not necessarily strongly depend on the difference in individuals because human beings can recognize a difference in individual movement in a few observation and structure some regularity regarding on the gesture quite naturally. The acquisition of regularity is very important in the development of human-friendly robot. Because robots must recognize the significance of human gesture without depending on how many gestures the robot observed.

Gestures involve speed, direction, the range of movement, etc. Focusing on this, we have proposed a method using Singular Value Decomposition to recognize gestures. As human beings, we evaluate similarity regardless of observed data length with Singular Value Decomposition. In the comparing correlation coefficients, Principal Component Analysis (PCA) and our proposed method, the accuracy of gesture recognition by our proposed method was higher than that of other methods, demonstrating it usefulness.

In a comparison between gesture distance method and gesture vector method, it is necessary, for the gesture vector method to prepare the number of elements of left singular vectors and to decide how to replace an element in left singular vectors calculated from checking data with elements using training data. For these reasons, the gesture vector method is not more practical than gesture distance method at present. For gesture distance method, recognition accuracy must be improved more. As shown in **Tables 1** to **3** and **5**, however, recognition results based on S_2 and S_5 showed a high recognition rate among several similarity measures. Since the formulation of S_2 and S_5 are the same, the total absolute differential of left singular vectors at the same order is suitable as a similarity definition for gesture recognition.

This result is very important in constructing human-friendly robots. In other words, our proposed method which calculates the total absolute differential of a left singular vectors at the same order is significant in evaluation because left singular vectors expresses a time-bin weight for identifying whole time-series movement. We thus find significance in expanding this method to this formulation when constructing knowledge in feature space is proposed. Regarding incorrect recognition, although the motion was incorrectly recognized as a gesture different from the intended one, motion data was quite similar. Gestures GR and GL, for example, have completely different meanings but their motions are very similar in that the hand waves left and right. Their difference lies in whether the hand moves faster from left to right or from right to left. Sometimes even human beings mistake distinguishing between them. We also need more discussion about how to distinguish knowledge in feature space and how to detect a folding point in continuous gestures.

6. Conclusions

Developing human-friendly robots makes the acquisition of regularity very important, because the robot must recognize the significance of human gesture without depending on how many gestures are observed. We therefore identified regularity using Singular Value Decomposition in movement recognition.

A 3D motion analysis algorithm using Singular Value Decomposition proposed for gesture recognition has been applied to gesture recognition. Results of experiments have showed the effectiveness of the algorithm. The algorithm will be improved and more recognition experiment discussed in future work.

Acknowledgements

This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology of Japan under Strategic Project to Support the Formation of Research Bases at Private Universities 2008-2012, in addition the Organization for Research and Development of Innovative Science and Technology (ORDIST) of Kansai University.

References:

[1] D. Tsuda, S. Wang, N. Miura, and Y. Jiang, "Development and Safety Validation of Lead for a Guidance Robot," J. of Biomedical Fuzzy Systems Association, Vol.11, No.1, pp. 43-48, 2009. (in Japanese)

[2] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review," Computer Vision and Image Understanding, Vol.73, No.3, pp. 428-440, 1999.

[3] A. Kendon, "Gesture: Visible Action as Utterance," Cambridge University Press, 2004.

[4] R. Bowden, "Learning Statistical Models of Human Motion," IEEE Workshop on Human Modelling, Analysis and Synthesis (CVPR2000), pp. 10-17, 2000.

[5] K. Furukawa, S. Igarashi, K. Ueno, T. Ozaki, S. Morita, N. Tamagawa, T. Okuyama, and I. Kobayashi, "Modeling Human Skill in Bayesian Network," Electric Transaction of Artificial Intelligence (ETAI), Linkoping University Electronic Press, 2002.

[6] S. Igarashi, K. Ueno, T. Ozaki, S. Morita, and K. Furukawa, "Skill Modeling in Cello Performance by Bayesian Networks," Technical Report of IEICE, Vol.102, No.709, pp. 1-6, 2003.

[7] R. Balasubramaniam and M. T. Turvey, "Coordination Modes in the Multisegmental Dynamics of Hula Hooping," Biological Cybernetics, Vol.90, pp. 176-190, 2004.

[8] S. Furuya and H. Kinoshita, "Organization of the Upper Limb Movement for Piano Key-depression Differs between Expert Pianists and Novice Players," Experimental Brain Research, Vol.185, No.4, pp. 581-593, 2008.

[9] S. Furuya, R. Osu, and H. Kinoshita, "Effective Utilization of Gravity During Arm Downswing in Keystroke by Expert pianists," Neuroscience, Vol.164, No.2, pp. 822-831, 2009.

[10] Y. Jiang, I. Hayashi, M. Hara, and S. Wang, "Three-dimensional Motion Analysis for Gesture Recognition Using Singular Value Decomposition," Proc of 2010 IEEE Int. Conf. on Information and Automation, pp. 805-810, 2010.

[11] I. Hayashi, Y. Jiang, M. Hara, and S. Wang, "Knowledge Acquisition from Motion of Evacuation Instruction Using Singular Value Decomposition," Proc. of the 26th Fuzzy System Symposium, pp. 824-829, 2010. (in Japanese)

[12] S. Mitra and T. Acharya, "Gesture recognition: a survey," IEEE Trans. on Systems, Man, and Cybernetics, Part C, Vol.37, No.3, pp. 311-324, 2007.

[13] Y. Iwai, K. Watanabe, Y. Yagi, and M. Yachida, "Gesture Recognition Using Colored Gloves," Proc. of IEEE Int. Conf. on Patter Recognition (ICPR96), Vol.A, pp. 662-666, 1996.

[14] M. V. Lamar, M. S. Bhuiyan, and A. Iwata, "Hand Gesture Recognition Using T-CombNET: A New Neural Network Model," IEICE trans. on information and systems, Vol.E83-D, No.11, pp. 1986-1995, 2000.

[15] T. E. Jerde, J. F. Soechting, and M. Flanders, "Biological Constraints Simplify the Recognition of Hand Shapes," IEEE Trans. on Biomedical Engineering, Vol.50, No.2, pp. 265-269, 2003.

[16] H. I. Suk, B. K. Sin, and S. W. Lee, "Hand Gesture Recognition Based on Dynamic Bayesian Network Framework," Pattern Recognition, Vol.43, pp. 3059-3072, 2010.

[17] C. Uras and A. Verri, "Hand gesture recognition from edge maps," Proc. of Int. Workshop on Automatic Face and Gesture Recognition, pp. 116-121, 1995.

[18] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A Real-Time Hand Gesture Recognition Method," Proc. of Int. Conf. on Multimedia and Expo (ICME2007), pp. 995-998, 2007.

[19] I. Hayashi, Y. Jiang, and S. Wang, "Embodied Knowledge of Gesture Motion Acquired by Singular Spectrum Analysis," Proc. of the First Int. Conf. on Vulnerability and Risk Analysis and Management (ICVRAM2010) and the Fifth Int. Symp. on Uncertainty Modeling and Analysis (ISUMA2010), pp. 45-52, 2011.

[20] T. Ide and K. Inoue, "Knowledge Discovery from Heterogeneous Dynamic Systems Using Change-point Correlations," Proc. 2005 SIAM Int. Conf. on Data Mining (SDM05), pp. 571-576, 2005.

[21] K. Mishima, S. Kanata, H. Nakanishi, Y. Horiguchi, and T. Sawaragi, "Extraction of Similarities and Differences in Human Behavior Using Singular Value Decomposition," Proc of the 53rd annual conf. of the Institute of Systems, Control and Information Engineers (SCI09), pp. 409-410, 2009. (in Japanese)

[22] M. E. Wall, A. Rechtsteiner, and L. M.Rocha, "Singular Value Decomposition and Principal Component Analysis." in A Practical Approach to Microarray Data Analysis, D. P. Berrar, W. Dubitzky, and M. Granzow (Eds.) Kluwer pp. 91-109, 2003.

[23] K. Doya, "Introduction to Computational Neuroscience: Toward Understanding of the Mechanisms of Learning in the Brain," Science Sha, 2007. (in Japanese)

[24] M. Kawato, "Internal Models for Motor Control and Trajectory Planning," Current Opinion in Neurobiology, Vol.9, Vol.6, pp. 718-727, 1999.

[25] R. C. Miall, D. J. Weir, D. M. Wolpert, and J. F. Stein, "Is the cerebellum a Smith predictor?" J. of Motor Behavior, Vol.25, pp. 203-216, 1993.

Name:
Isao Hayashi

Affiliation:
Professor, Graduate School of Informatics, Kansai University

Address:
2-1-1 Ryozenji-cho, Takatsuki, Osaka 569-1095, Japan

Brief Biographical History:
1981-1983 Sharp Corporation
1987-1993 Panasonic Corporation
1993-2004 Hannan University
2004- Kansai University

Main Works:
• "Vitroid – The Robot System with an Interface Between a Living Neuronal Network and Outer World," Int. J. of Mechatronics and Manufacturing System, Vol.4, No.2, pp. 135-149, 2011.

Membership in Academic Societies:
• International Fuzzy Systems Association (IFSA)
• The Institute of Electrical and Electronics Engineers, INC. (IEEE)
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• Japanese Neural Network Society (JNNS)

Name:
Yinlai Jiang

Affiliation:
Department of Intelligent Mechanical Systems Engineering, Kochi University of Technology

Address:
185 Miyanokuti, Tosayamada, Kami, Kochi 782-8502, Japan

Brief Biographical History:
2008 Recieved Ph.D. in Engineering, Kochi University of Technology
2008- Kochi University of Technology

Main Works:
• Y. L. Jiang, S. Y. Wang, K. Ishida, T. Ando, and M. G. Fujie, "Directional intention identification for running control of an omni-directional walker," J. of Advanced Computational Intelligence and Intelligent Informatics, Vol.14, No.7, pp. 784-792, 2010.

Membership in Academic Societies:
• The Institute of Electrical and Electronics Engineers, INC. (IEEE)
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• The Robotics Society of Japan (RSJ)
• Japan Society of Medical and Biological Engineering (JSMBE)

Name:
Shuoyu Wang

Affiliation:
Department of Intelligent Mechanical Systems Engineering, Kochi University of Technology

Address:
185 Miyanokuti, Tosayamada, Kami, Kochi 782-8502, Japan

Brief Biographical History:
1993 Received Ph.D. in Electrical Engineering, Hokkaido University
1993-1997 Yamagata University
1997- Kochi University of Technology

Main Works:
• S. Y. Wang, T. Tsuchiya, and M. Mizumoto, "Distance-Type Fuzzy Reasoning Method, Journal of Biomedical Fuzzy Systems Association," Vol.1, No.1, pp. 61-78, 1999.

Membership in Academic Societies:
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• The Robotics Society of Japan (RSJ)
• The Japan Society of Mechanical Engineering (JSME)
• The Society of Instrument and Control Engineers (SICE)
• Japanese Society for Medical and Biological Engineering (JSMBE)
• The Society of Life Support Engineering (LSE)