Analysis and Extraction of Knowledge from Body Motion Using Singular Value Decomposition

Yinlai Jiang, Isao Hayashi and Shuoyu Wang

Abstract—The dexterity of body motion when performing skills are being actively studied. In this paper, singular value decomposition is used to extract the dexterous features from the time-series data of body motion. A matrix is composed by overlapping the subsets of the time-series data. The left singular vectors of the matrix are extracted as the patterns of the motion and the singular values as a scalar, by which each corresponding left singular vector affects the matrix. A gesture recognition experiment, in which we categorize gesture motions with indexes of similarity and estimation that use left singular vectors, was conducted to validate the method. Furthermore, in order to understand the features better, the features of the left singular vectors were described as fuzzy sets, and fuzzy if-then rules were used to represent the knowledge.

I. INTRODUCTION

Skills remembered by the human body and reflected by the dexterity of body motion is called embodied knowledge which plays an important role in acquiring skills. Embodied knowledge is a native human endowment that is usually not consciously accessible.

Embodied knowledge has been studied by cross-disciplinary approaches. Kudo et al. studied the process of motor skill acquisition and proposed a dynamical systems approach to understand the principle underlying organization of experts’ and novices’ movements [1]. Miall et al. proposed a forward model to output a signal to a controlled object via feedback control using the Smith predictor having no time delay for the control [2]. Kawato proposed a cerebellar computational model, called an internal model, which argued that the inverse model with feedback and feedforward control is useful in modeling motor control [3], [4]. 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. Our approach assumes that skills consist of a hierarchical structure with a mono-functional layer to generate the single function result and a meta-functional layer that adapts itself to environmental change. We assume that a skill is the knowledge embodied in the human body.

The information of motion, such as trajectory, speed, and acceleration, has been measured to extract and analyze embodied knowledge [5], [6], [7], [8], [9], [10]. Time-series data analysis is usually necessary to extract features from the measurement data. A variety of methods for the analysis of time-series data of physical movement have been proposed. Lamar et al. [11] proposed a neural network, Temporal-CombiNET (T-CombiNET), and applied it to Japanese-Kana finger spelling recognition. Suk et al. [12] recognized hand gestures in a continuous video stream by a dynamic Bayesian network. Jerde et al. [13] measured the angles of hand joints for recognizing fingerspelling hand shapes, and reduced the dimensions of the hand by discriminant analysis by Principal Component Analysis (PCA). Daffertshofer et al. [14] also suggested the effectiveness of PCA for reducing high-dimensional time-series data sets to a small number of modes in the analysis of gait kinematics. Mitra et al. [15] surveyed methods of gesture recognition and suggested that the Hidden Markov model is effective for gesture motion data analysis.

However, since a neural network is too sensitive in time-series data length, the accuracy is not very good. PCA reduces the number of explanatory variables and is a 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 data is inadequate. In other words, the accuracy of PCA declines when the contribution ratio is low due to a shortage of data. The Hidden Markov models are not effective when the number of states is large or the data is discontinuous.

Recently, singular value decomposition (SVD) [16], [17] has been used in time-series data analyses for data mining [18] and motion analyses to study the coordinative structures in human behavior [19]. In this paper, a new method to extract embodied knowledge of body motion from time-series data by using SVD is proposed. In our method, the left and right singular vectors and the singular values are decomposed from a Hankel matrix defined from the time-series data, which are measured with sensors [20], [21]. Since the left singular vector represents the characteristics of the Hankel matrix and the singular value represents the strength of the corresponding left singular vector, SVD is used more generally as a method for extracting characteristics from observed time-series data.

The remainder of this paper is organized as follows. Section II introduces the method to extract features from time-series data of motions. In Section III, the proposed method is applied to a hand gesture recognition experiment, in which we distinguish five kinds of gestures according to the similarity
and the estimation by using the left singular vectors. The characteristic and significance of the embodied knowledge extraction using SVD is discussed based on the results. Section IV proposes a method to represent the knowledge in the left singular vectors with fuzzy if-then rules. Finally, Section VI concludes the paper and discusses the future work.

II. EMBODIED KNOWLEDGE EXTRACTION USING SVD

Body motion is usually measured by multiple sensors. Suppose that w points \((P_1, P_2, \ldots, P_w)\) of the body are measured while a person is performing a motion. On point \(P_i\), the measured data series of motion \(G\) is denoted as \(\tau_i.G\). The data series of \(\tau_i.G\) consists of three-dimensional data \((X_i.G, Y_i.G, Z_i.G)\). From the time-series data \(\tau_i.G = (X_i.G, Y_i.G, Z_i.G)\), \(n\) vectors by \(m\) data sampling are extracted by overlapping and the matrices \(M_X^i.G, M_Y^i.G\), and \(M_Z^i.G\) are constructed as a collective of the measurement data on the \(X, Y,\) and \(Z\) coordinates of the motion, respectively. Fig. 1 shows a design for constructing the matrix \(M_X^i.G\). The matrices \(M_X^i.G, M_Y^i.G,\) and \(M_Z^i.G\) are described as follows:

\[
\begin{align*}
M_X^{i,G} &= (X_1^{i,G}, X_2^{i,G}, \ldots, X_n^{i,G})^T \\
M_Y^{i,G} &= (Y_1^{i,G}, Y_2^{i,G}, \ldots, Y_n^{i,G})^T \\
M_Z^{i,G} &= (Z_1^{i,G}, Z_2^{i,G}, \ldots, Z_n^{i,G})^T
\end{align*}
\]

where \(X_p^{i,G} = (x_{p,1}^{i,G}, x_{p,2}^{i,G}, \ldots, x_{p,n}^{i,G})\), \(p = 1, 2, \ldots, n\), and \(x\) is a datum on the \(X\) coordinate. We define \(Y_p^{i,G}\) and \(Z_p^{i,G}\) in the same way. Since the matrix is designed by overlapping the extracted data from the whole time-series data, the method using SVD is less constrained by the length of the whole data than those using a neural network. The difference between PCA and our method is that PCA generally analyzes the matrix composed of the deviation of each datum from the empirical mean and does not allow overlap of the measured data.

The SVD is an important factorization of a rectangular real or complex matrix with many applications in signal processing and statistics. Applications using SVD include computing the pseudo inverse, least squares data fitting, matrix approximation, and determining the rank, range, and null space of a matrix [16], [17]. Suppose \(M\) is an \(m\)-by-\(n\) matrix, then a factorization of \(M\) is \(M = USV^T\), where \(U = (u_1, u_2, \ldots, u_m)\) contains the left singular vectors of \(M\), \(V = (v_1, v_2, \ldots, v_n)\) contains the right singular vectors of \(M\), and the matrix \(\Sigma\) is an \(m\)-by-\(n\) diagonal matrix with nonnegative real singular values on the diagonal.

Suppose \(M_k^{i,G}, k = \{X, Y, Z\}\) is an \(m\)-by-\(n\) matrix as the general format of \(M_X^{i,G}, M_Y^{i,G}, M_Z^{i,G}\), which is composed by overlapping the subsets of measurement data. The SVD of the matrix \(M_k^{i,G}\) is

\[
M_k^{i,G} = U_k^{i,G}S_k^{i,G}V_k^{i,G}^T
\]

where \(U_k^{i,G} = (u_1^{i,k}, u_2^{i,k}, \ldots, u_m^{i,k})\) is an \(m\)-by-\(n\) unitary matrix, \(V_k^{i,G} = (v_1, v_2, \ldots, v_n)\) is an \(n\)-by-\(n\) unitary matrix, and the matrix \(\Sigma_k^{i,G}\) is an \(m\)-by-\(n\) diagonal matrix. The diagonal entries of \(\Sigma_k^{i,G}\) are the singular values of \(M_k^{i,G}\). The matrix \(U_k^{i,G}\) contains the left singular vectors of \(M_k^{i,G}\) and the matrix \(V_k^{i,G}\) contains the right singular vectors of \(M_k^{i,G}\).

Now, take \(M_X^{i,G}\) as an example of matrix \(M_k^{i,G}\) to discuss motion analysis. SVD can decompose the matrix \(M_X^{i,G}\) into a product of \(U_X^{i,G}, S_X^{i,G}\), and \(V_X^{i,G}\). Intuitively, the left singular vectors in \(U_X^{i,G}\) form a set of patterns of \(M_X^{i,G}\) and the diagonal values in matrix \(S_X^{i,G}\) are the singular values, which can be considered as scalars, by which each corresponding left singular vectors affect the matrix \(M_X^{i,G}\). Suppose that the number of left singular vectors is \(l\) and the element number of the \(j\) th left singular vector is \(q\). Let us denote the couples of the singular values and the left singular vector as \(((\sigma_1^j, u_1^{j,1}), (\sigma_2^j, u_2^{j,2}), \ldots, (\sigma_l^j, u_l^{j,l}))\) for \(u_j^{i,G} = (\bar{u}_{j,1}^{i,G}, \bar{u}_{j,2}^{i,G}, \ldots, \bar{u}_{j,q}^{i,G})\) in the descending order of the singular values, where \(\bar{u}_{h,1}^{i,G}\) is the \(h\) th element of the \(j\) th left singular vector \(u_j^{i,G}\). The left singular vector expresses the characteristic of the whole time-series data better if its corresponding singular value singular is larger. That is, the greater the singular value is, the more dominant the corresponding pattern is.

III. GESTURE RECOGNITION WITH LEFT SINGULAR VECTORS

Since the left singular vectors \((u_1^{i,G}, u_2^{i,G}, \ldots, u_l^{i,G})\) well represent the characteristics of motions, they can be utilized as features to recognize and classify the motions. We conducted a gesture recognition experiment to demonstrate the effectiveness of feature extraction using the left singular vectors.

A. Motion Measurement of Hand Gesture

Five kinds of hand gestures, CH (Come here), GA (Go away), GR (Go right), GL (Go left), and CD (Calm down) which are commonly used in daily life, were performed by two subjects, SW and ST. The gestures were performed in a \(50cm \times 50cm \times 50cm\) cubic space, whose zero point and
coordinate system are shown in Fig. 2. The motions of the hand gestures were measured with Movetr/3D and GE60/W (Library, Tokyo, Japan). The subjects were instructed to finish the gestures within the same speed period. Five markers were measured: $P_1$ on the tip of the thumb, $P_2$ on the tip of the middle finger, $P_3$ on the tip of the little finger, $P_4$ on the thumb side of the wrist and $P_5$ on the little finger side of the wrist.

The motion of subject SW performing CH is shown in Figure 3 by nine frames extracted from the experimental video every $1/6$ s. The measurement time-series data of $P_2$ when subject SW performed the five kinds of gestures are shown in Figure 4. In Figure 4, the movement change for GA, CH, and CD is large in the top and bottom direction (onto the $z$-axis) and in the front and back direction (onto the $y$-axis), and for GR and GL, the movement change is big in the right and left direction (onto the $x$-axis). One gesture was executed 9 times by each subject to get a sufficient variety of motion. Data of the first five executions were used as the acquisition of patterns of the gesture. Data of the last four times were used to distinguish the gesture.

We proposed a gesture recognition method using the left singular vectors extracted from the time-series data of gesture motion based on the similarity between gesture distances (SGD). Our aim is to discuss a method to extract embodied knowledge from the time-series data rather than to develop a method to recognize hand gestures.

**B. Gesture Recognition Based on SGD**

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

![Fig. 2. Environment for Measurement of Gestures](image)

![Fig. 3. Motion of Gestures(Representative frames are extracted every 1/6 second)](image)
We define the left singular vectors $\tau_{X,CHD}$ for checking data as $U_{X,CHD} = (u_{1,X,CHD}^j, u_{2,X,CHD}^j, \ldots, u_{l,X,CHD}^j)$, for $u_{j,X,CHD} = (u_{i,j,X,CHD}, u_{i,j,X,CHD}, \ldots, u_{i,j,X,CHD})$ in the same way.

To study the usage of the left singular vectors, three kinds of similarity criteria to recognize the hand gestures are defined on the data $\tau_{i,G} = (X^{i,G}, Y^{i,G}, Z^{i,G})$ as follows:

$$S_1 : r_i(U_{TRD}^G, U_{CHD}^j) = \frac{1}{3lq} \sum_{k=1}^{3} \sum_{j=1}^{l} \sum_{h=1}^{q} |u_{i,j,k,TRD} - u_{i,j,k,CHD}|$$  

$$S_2 : r_i(U_{TRD}^G, U_{CHD}^j) = \frac{1}{3lq} \sum_{k=1}^{3} \sum_{j=1}^{l} \sum_{h=1}^{q} |u_{i,j,k,TRD} - u_{i,j,k,CHD}|$$  

$$S_3 : r_i(U_{TRD}^G, U_{CHD}^j) = \frac{1}{3lq} \sum_{k=1}^{3} \sum_{j=1}^{l} \sum_{h=1}^{q} (u_{i,j,k,TRD} - u_{i,j,k,CHD})^2$$

The similarity $S_1$ is defined by the absolute differential of the total left singular vector between the training data and the checking data. The similarity $S_2$ is defined by the total absolute differential of the left singular vectors between the training data and the checking data at the same order. The similarity $S_3$ is defined by the Euclidean distance between the left singular vectors of the training data and the checking data on the multidimensional space.

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

$$E_1 : G^* = \{G_f | \max_i \sum_{j} n(G_f^j) \}$$

$$E_2 : G^* = \{G_f | \min_i \sum_{j} r_i(U_{TRD}^G, U_{CHD}^j) \}$$

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

The estimation $E_1$ is defined by counting the number of minimal similarity values on all the markers and the most counted gesture is output as the recognition result. The estimation $E_2$ outputs the gesture with the minimal total similarity values as the recognition result.

**C. Gesture Recognition Results**

Table I shows the recognition results of the gesture patterns of two subjects based on the three kinds of different similarities. Equation (5), Equation (6), and Equation (7). In the calculation, $m = 125$, $n = 5$, $q = 125$, $l = 1$, $w = 5$. The recognition results suggest that similarities $S_2$ and $S_3$ lead to relatively higher recognition accuracy. Especially, the pair of the similarity $S_2$ and the estimation $E_1$ leads to a significantly high accuracy of 90.0%. The result suggests that the pair of $S_2$ and $E_1$ is more feasible for gesture recognition. Regarding the incorrect recognitions, motion data was quite similar although the motion was incorrectly recognized as a gesture different from the intended one. Gestures GR and GL, for example, have the opposite meanings but their motions are very similar in that the hand waves left and right. Their difference lies in that the hand moves faster from left to right in GR while faster from right to left in GL. Even human beings sometimes mistake distinguishing between them.

The recognition results of the two method show that the left singular vectors extracted from the time-series data can be used as knowledge to distinguish gestures. Especially, the total absolute differential of a left singular vectors at the same order is significantly effective because the left singular vectors expresses a time-dependent weight for identifying the whole movement.

In the experiment, the positions of the five markers on the right hand were measured. However, it is possible that not all the positions of these markers have a high relevance.
to the gestures. We compared the accuracy of each marker for gesture recognition with similarity $S_2$ and estimation $E_1$ by calculating the left singular vectors at each marker. Table II shows the large counting number of minimized similarity values. The results are shown in the form of $a/b$, where $a$ is the number of counted times of the most counted gesture and $b$ is the name of the gesture. As a result, the first marker $P_1$ was selected as the most important marker because the accuracy is the highest, 93.85%. Since $P_1$ measures the time series at the tip of the thumb and is largely related to movement of the thumb, it is suggested that the motion of the thumb is important in gesture recognition. Reducing the number of markers considered in the recognition not only reduces the calculation but also improves the recognition accuracy.

Since the method is less affected by the length of the time-series data, it leads to a high degree of accuracy of hand gesture recognition. The recognition obtained the best accuracy with the similarity calculation of $S_2$. The similarity $S_2$ is defined by the total absolute differential of the left singular vectors between the training data and the checking data at the same order. It is suggested that the order of the time-series data is an important characteristic of the time-series data and the left singular vector is able to extract that characteristic. As for the estimation methods, $E_1$, which is defined by counting the number of minimal similarity values on all the markers, obtained better accuracy. Motion analysis usually deals with data from multiple sensors and the sensors are related to each other; for example, $P_4$ and $P_5$ in the gesture motion measurement, which are on the thumb side of the wrist and on the little finger side of the wrist, respectively, are highly correlated. In this case, a decision by the majority is more effective than averaging or summing.

### IV. Knowledge Representation with Fuzzy If-Then Rules

The results of previous section illustrate that the left singular vectors contain the features of gesture motions. The features can be used to classify the gestures with a high degree of accuracy by using a proper calculation. In order to understand the features better, an intuitive and understandable way to represent these features are required. In this section, the features of the left singular vectors are described as fuzzy sets, and fuzzy if-then rules [24] are used to represent the knowledge.

The fuzzy if-then rule to represent the knowledge assumes the following form

$$\text{if } \bigwedge_{i=1}^{w} \bigwedge_{j=1}^{l} (Q_{i,k}^{j}G) \text{ is } Q_{i,k}^{j}G \text{ then } G^* \text{ is } G$$

(10)

where $Q$ is a quadrilateral fuzzy set to represent the $j$th singular vector on the $k$th coordinate at the $i$th marker. $Q$ is determined by the frequency distribution of the elements in the left singular vectors.

The histograms, which show the frequency distribution of the elements of the 1st - 3rd left singular vectors, are shown in Figure 5 as examples. The left vertical axis indicates the frequency. The singular vectors were calculated from $M_X$, that is, the matrix composed from the x coordinate time-series data at point 1 of gesture CH. The elements are between -0.4 and 0.4, and are counted in 4 intervals: (-0.4, 0.2), [0.2, 0), [0, 0.2), and [0, 0.4). The elements are between -0.4 and 0.4, and are counted in 4 intervals: (-0.4, 0.2), [0.2, 0), [0, 0.2), and [0.2, 0.4). The quadrilateral fuzzy set to describe the features is denoted as $Q(a_1, a_2, a_3, a_4)$, where $a_1 \sim a_4$ are the lateral coordinates, and $d_1 \sim d_4$ the vertical coordinates of the vertices of the quadrilateral. $d_1 \sim d_4$ are the grade of $a_1 \sim a_4$. $a_1 \sim a_4$ and $d_1 \sim d_4$ are determined to make the quadrilateral fuzzy set to approximate the shape of the histogram with $d_1$ and $d_4$ being 0, and the maximum of $d_2$ and $d_3$ being 1. Figure 5 shows the quadrilateral fuzzy sets of the corresponding histograms. The right vertical axis indicates the grade of membership. The fuzzy sets for the 1st-3rd left singular vectors are $Q(-0.32(0), -0.1(1), 0.1(0), 0.3(0))$, $Q(-0.32(0), -0.1(1), 0.1(0.77), 0.36(0))$, $Q(-0.42(0), -0.1(0.88), 0.1(1), 0.3(0))$, respectively. Since these fuzzy rules represent the knowledge about the features of the gestures, gesture recognition can be realized by fuzzy reasoning based on these fuzzy rules.

### V. Conclusions

We propose to extract embodied knowledge by identifying the relationship between the sensor input and the output of the internal model. The mechanism underlying the skillful movement which a human being performs unconsciously is still not sufficiently elucidated from the viewpoints of neurophysiology or the structure of the body. The motion analysis method using SVD we propose is effective to extract...
embodyed knowledge from motion by obtaining the characteristics and their strength. Its effectiveness was validated by the gesture recognition experiment. In addition, fuzzy if-then rules were used to represent the extracted knowledge. Motion analysis using SVD is suggested as a useful method to extract embodied knowledge from motion measurement data. Future work will develop a motion recognition method with fuzzy reasoning based on the fuzzy if-then rules extracted.

REFERENCES


