Embodied Knowledge of Gesture Motion Acquired by Singular Spectrum Analysis

I. Hayashi[1], Y. Jiang[2], and S. Wang[2]

1. Faculty of Informatics, Kansai University, 2-1-1, Ryozenji-cho, Takatsuki, Osaka 569-1095, Japan; PH (072) 690-2448; FAX (072) 690-2491; email: ihaya@cbii.kutc.kansai-u.ac.jp
2. School of Systems Engineering, Kochi University of Technology, 185 Miyanokuti, Tosayamada, Kami, Kochi 782-8502, Japan; PH (088) 757-2013; FAX (088) 757-2013; email: {jiang.yinlai, wang.shuoyu}@kochi-tech.ac.jp

ABSTRACT

Whenever a disaster occurs, it's of utmost importance that the rescue system recognizes accurately human behavior and evacuation command in the fire and its black smoke. However, we have infinite pattern for movement instructions by our personality. On the other hand, the singular spectrum analysis method has proposed as analytical method for time-series data. In this paper, we propose a method for acquiring embodied knowledge of human behavior from time-series gesture data using singular spectrum analysis. A behavior is distinguished in terms of gesture characteristic with similarity criteria by interval time-series data. We discuss the usefulness of the proposed method using an example of gesture motion.

I. INTRODUCTION

In order to assure safety and security in the occurrence of a natural disaster or a large-scale accident, it is important to communicate with each other in the disaster, and such how to communicate can allow us to detect and avoid more dangers. For the communication tool, we need a development of monitoring system which records human behavior in the disaster, and distinguishes the gesture motion and informs people a safety escape route automatically adding safety intelligence, e.g., disaster information, criminal information. Especially, it's of utmost importance for evacuees to recognize commands of inducer accurately to find an escape route under fire and black smoke in the disaster. However, it is difficult to recognize an evacuation command from inducer’s infinite gesture motion, and so we need a system which can recognize human behavior automatically [1].

In this paper, we aim a development of rescue robot which sense a movement of evacuation command and acquire embodied knowledge of the movement [2, 3]. A gesture recognition method [4] is proposed to enable the rescue robot to communicate with humans. Gesture recognition has been studied extensively and there have been varied approaches to handle gesture recognition, ranging from mathematical models based on hidden Markov chains to tools or approaches based on soft computing [5]. We discuss a new gesture recognition method to identify 3-dimensional gesture motions using singular value decomposition (SVD). Applications which employ the SVD include computing the pseudoinverse, least squares fitting of data, matrix approximation, and determining the rank, range and null space of a matrix [6]. Recently, the SVD have been utilized in time-series data analysis for knowledge discovery [7] and motion analysis to extract similarities and differences in human behavior [8]. In our proposed model, we measure the similarity criteria between the gesture of evacuation command and the instruction we learned before using left singular vectors and singular values decomposition, and distinguish the gestures. We proposed two kinds of methods, first method to measure the similarity between the gesture distances and the second method to measure the similarity of the gesture vector. We discuss the usefulness of the proposed methods using an example of five kinds of 3-dimensional gesture motions.

II. MESUREMENT OF 3-DIMENSIONAL EVACUATION GESTURE

The motions of the hand gestures are measured with Movetr/3D and GE60/W (Library, Tokyo, Japan). Subjects are two males, SW and ST, in twenties. Five markers, M_1 on the tip of the thumb, M_2 on the tip of the middle finger, M_3 on the tip of the little finger, M_4 on the thumb-side of the wrist and M_5 on the little finger side of the wrist, were measured. The gestures were performed in a 50cm×50cm×50cm cubic space shown in Figure 1. In the experiment, five kinds of hand gestures, CH (Come here), GA (Go away), GR (Go right), GL (Go left), and CD (Calm down), were performed by subjects. One gesture was executed 9 times by each subject. Data of the first 5 times execution were used as patterns of the gesture. Data of last 4 times were used to be distinguished.

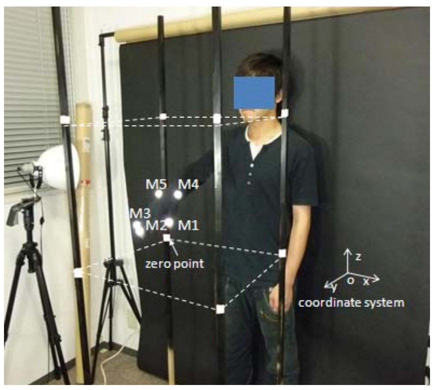

Figure 1: Experiment

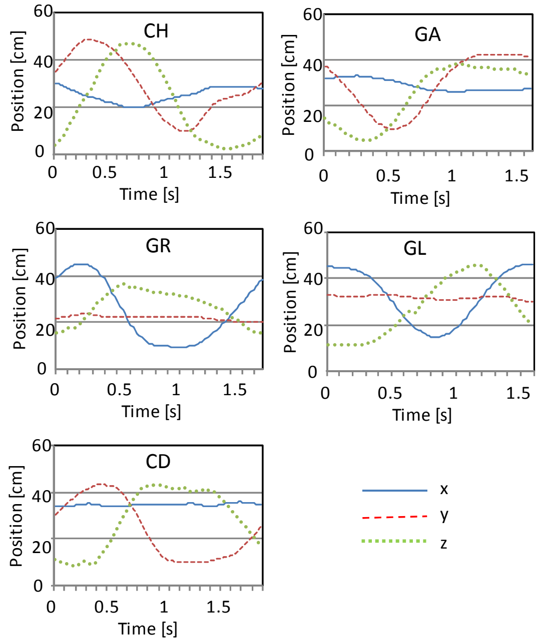

Figure 2: Gesture of Examples (M2 of SW)

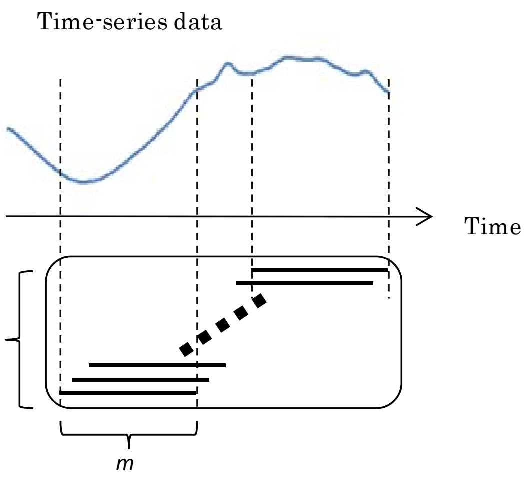

Figure 3: Design of Matrix $M_X^{i,G}$

The measurement time-series data of M_2 when subject SW performed the five kinds of gestures are shown in Figure 2. A movement change as for GA, CH, and CD is big in the top and bottom direction (onto *z*-axis) and in the front and back direction (onto *y*-axis), and as for GR and GL, the movement change is big in the right and left direction (onto *x*-axis).

III. GESTURE ANALYSIS USING SINGULAR VALUE DECOMPOSITION

Suppose *M* is an m-by-n matrix. Then there exists a factorization of the form: $M = U \Sigma V^T$, where $U = (u_1, u_2, \ldots, u_m)$, $V = (v_1, v_2, \ldots, v_n)$, and the matrix Σ is *m*-by-*n* diagonal matrix with nonnegative real numbers on the diagonal. The matrix *U* contains the left singular vectors of *M* and the matrix *V* contains the right singular vectors of *M*. Suppose that there are *w* measurement points (P_1, P_2, ..., P_w). On point P_i, the measured data series of gesture *G* is denoted as $\tau^{i,R}$, which consists of 3-dimentional data ($X^{i,G}$, $Y^{i,G}$, $Z^{i,G}$). We detect the time series $X_X^{i,G} = (x_1^{i,G}, x_2^{i,G}, \ldots, x_n^{i,G})^T$ contains the *x* coordinate values of the P_i point. Then matrix $M_X^{i,G}$ is defined as a collective of the change of *x* coordinate values of the gesture, $M_X^{i,G} = (X_1^{i,G}, X_2^{i,G}, \ldots, X_n^{i,G})$.

The matrix $M_X^{i,G}$ can be decomposed into a product of $U_X^{i,G}$, $\Sigma_X^{i,G}$ and $V_X^{i,G}$. The design of matrix $M_X^{i,G}$ is shown in Figure 3. Let us denote the singular values and the left singular vectors as $((\delta_{1,X}^{i,G}, u_{1,X}^{i,G}), (\delta_{2,X}^{i,G}, u_{2,X}^{i,G}), \cdots, (\delta_{l,X}^{i,G}, u_{l,X}^{i,G}))$, for $u_{j,X}^{i,G} = (u_{1j,X}^{i,G}, u_{2j,X}^{i,G}, \cdots, u_{hj,X}^{i,G}, \cdots, u_{qj,X}^{i,G})$ in descending order of the singular values. The parameter *l* represents the number of representative patterns under consideration, and the parameter *q* represents the number of elements of the singular vector. The left singular vectors, $u_{1,X}^{i,G}$, $u_{2,X}^{i,G}$, ... , $u_{l,X}^{i,G}$ of $M_X^{i,G}$, represent the change patterns of the *x* coordinate values on this point of the hand gesture. We proposed two kinds of motion analysis methods for gesture recognition using SVD.

1) ***Method for Similarity between Gesture Distances***

Suppose that the measured data series are divided into $\tau_{TRD}^{i,G}$ as reference data series and $\tau_{CHD}^{i,G}$ as data series to be recognized. Let us denote the left singular vectors of $\tau_{X,TRD}^{i,G}$ related to the *x* coordinate values of the P_i point on the hand while a hand 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 $U_{j,X,TRD}^{i,G} = (u_{1j,X,TRD}^{i,G}, u_{2j,X,TRD}^{i,G}, \cdots, u_{hj,X,TRD}^{i,G}, \cdots, u_{qj,X,TRD}^{i,G})$, and the left singular vectors of $\tau_{X,CHD}^{i}$ as $u_{X,CHD}^{i} = (u_{1,X,CHD}^{i}, u_{2,X,CHD}^{i}, \ldots, u_{l,X,CHD}^{i})$, for $u_{j,X,CHD}^{i,G} = (u_{1j,X,CHD}^{i,G}, u_{2j,X,CHD}^{i,G}$, . $\cdots, u_{hj,X,CHD}^{i,G}, \cdots, u_{qj,X,CHD}^{i,G})$. Three kinds of similarity criteria between gestures related to data series of 3-dimentional data ($X^{i,G}$, $Y^{i,G}$, $Z^{i,G}$) are defined as follows;

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

$$S_2: \; r_i(u_{TRD}^{i,G}, u_{CHD}^{i}) = \frac{1}{3lq} \sum_{k=1}^{3} \sum_{j=1}^{l} \sum_{h=1}^{q} \left| u_{hj,k,TRD}^{i,G} - u_{hj,k,CHD}^{i} \right| \tag{2}$$

$$S_3: \; r_i(u_{TRD}^{i,G}, u_{CHD}^{i}) = \frac{1}{3lq} \sum_{k=1}^{3} \sqrt{\sum_{j=1}^{l} \sum_{h=1}^{q} \left(u_{hj,k,TRD}^{i,G} - u_{hj,k,CHD}^{i} \right)^2} \, . \tag{3}$$

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

$$E_1:\ G^* = \left\{ G_f \mid \max_{\mathrm{f}} \sum_{i=1}^{w} n(G_f^i),\ \text{ for }\ G_f^i = \{G_f \mid \min_f r_i(u_{TRD}^{i,G_f}, u_{CHD}^i)\} \right\} \tag{4}$$

$$E_2:\ G^* = \left\{ G_f \middle| \min_f \sum_{i=1}^{w} r_i(u_{TRD}^{i,G_f}, u_{CHD}^i) \right\} \tag{5}$$

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

2) ***Method for Similarity between Gesture Vectors***

If one of the data series $X_p^{i,G}$ in $M_X^{i,G}$ is replaced by another data series X_{CHD}^i, the singular values and left singular vectors of $M_{X,CHD}^i$ will be different from those of $M_X^{i,G}$.

$$M_{X,CHD}^i = (X_1^{i,G},\ X_2^{i,G},\ \ldots,\ X_{p-1}^{i,G},\ X_{CHD}^i,\ X_{p+1}^{i,G},\ \ldots,\ X_n^{i,G}) \tag{6}$$

The difference between the left singular vectors of X $M_X^{i,G}$ and $M_{X,CHD}^i$ is determined by how X_{CHD}^i is different from the other p-1 data series. Therefore, if X_{CHD}^i comes from another kind of hand gesture, the difference can be utilized as a criterion for judging whether X_{CHD}^i comes from the same kind of hand gesture as the other data series. In our method, the location of X_{CHD}^i is fixed in the end of data series, and three kinds of similarity between gestures related to data series of 3-dimentional data ($X^{i,G}$, $Y^{i,G}$, $Z^{i,G}$) are defined as follows;

$$S_4:\ r_i(u_{TRD\backslash}^{i,G}, u_{CHD}^i) = \frac{1}{3lq} \sum_{k=1}^{3} \sum_{j=1}^{l} \left| \sum_{h=1}^{q} u_{hj,k,TRD}^{i,G} - \sum_{h=1}^{q} u_{hj,k,CHD}^{i} \right| \tag{7}$$

$$S_5: \ r_i(u_{TRD}^{i,G}, u_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sum_{j=1}^{l}\sum_{h=1}^{q}\left|u_{hj,k,TRD}^{i,G} - u_{hj,k,CHD}^{i}\right| \tag{8}$$

$$S_6: \ r_i(u_{TRD}^{i,G}, u_{CHD}^{i}) = \frac{1}{3lq}\sum_{k=1}^{3}\sqrt{\sum_{j=1}^{l}\sum_{h=1}^{q}\left(u_{hj,k,TRD}^{i,G} - u_{hj,k,CHD}^{i}\right)^2}\ . \tag{9}$$

Since there are w measurement points (P_1, P_2, ..., P_w), the estimated gesture G^* is identified by the following estimation;

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

IV. RESULTS AND DISCUSSIONS

In order to show the usefulness of the proposed methods, we distinguished gestures of two subjects, SW and ST using two kinds of methods. Since the average number of SW and ST's gestures is 125.2 (SD: 29.0), the number of data m is set to be 125. We also set n=5, q=125, l=1, and w=5.

1) *Method for Similarity between Gesture Distances*

Table I shows the recognition results. The pair of the similarity S_2 and the estimation E_1 is 90.0 %. The recognition results suggest that the pair of S_2 and E_1 is more feasible in gesture recognition. Table 2 and Table 3 show the counting number of measurement points for two subjects with the pair of S_2 and E_1. The recognition results suggest that the gestures of CH, GA, and CD are distinguished well, but it is not so well for two gestures of GR and GL of SW. However, in general it is hard to distinguish between a gesture of GR (Go right) and GL (Go left), and so the results are understandable.

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%

2) *Method for Similarity between Gesture Vectors*

Table 4 shows the recognition results based on the three kinds of similarity definitions of S_4, S_5, and S_6. The recognition results suggest that similarity definitions of S_5 and S_6 leaded to relatively higher correct recognition rates while the correct rate of the recognition based on S_4 was very low. Therefore, S_5 and S_6 are more feasible in gesture recognition.

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: Results of Method for Similarity between Gesture

Estimated Gesture	Similarity (S_4)		Similarity (S_5)		Similarity (S_6)	
	Correct	Others	Correct	Others	Correct	Others
SW_CH	1	3	2	2	1	3
SW_GA	3	1	4	0	4	0
SW_GR	2	2	3	1	3	1
SW_GL	1	3	3	1	3	1
SW_CD	0	4	3	1	3	1
ST_CH	2	2	2	2	2	2
ST_GA	1	3	4	0	4	0
ST_GR	3	1	4	0	4	0
ST_GL	0	4	3	1	2	2
ST_CD	0	4	4	0	4	0
Correct Rate	30.3%		80.0%		75.0%	

3) ***Discussions***

Similar to speech and handwriting, gestures vary between individuals, even for the same individual between different instances. However, as shown in Table 1 to Table 4, the recognition results based on S_2 and S_5 illustrated high recognition rate among several similarity measures. Since the formulation of S_2 and S_5 are same, the absolute differential of the left singular vectors at the same order is suitable for gesture recognition as similarity definition. As for the incorrect recognitions, for example, the gestures GR and GL 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 humans make mistakes in distinguishing them from each other.

V. CONCLUSIONS

In this paper, a novel 3D motion analysis algorithm using singular value decomposition (SVD) is proposed for gesture recognition. We applied the proposed method to gesture recognition, and the experiment results verified the effectiveness of the algorithm. This work was partially supported 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.

REFERENCES

[1] D.Tsuda, S.Wang, N.Miura, and Y.Jiang: Development and safety validation of lead for a guidance robot, *Journal of Biomedical Fuzzy Systems Association*, Vol.11, No.1, pp.43-48 (2009) [in Japanese].

[2] Y.Jiang, I.Hayashi, M.Hara, and S.Wang: Three-dimensional Motion Analysis for Gesture Recognition Using Singular Value Decomposition, *Proc of 2010 IEEE International Conference on Information and Automation*, pp.805-810 (2010).

[3] 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]

[4] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge (2004).

[5] S. Mitra and T. Acharya: Gesture recognition: a survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol.37, issue 3, pp.311- 324 (2007)

[6] 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, M.Granzow, eds*. pp.91-109, Kluwer: Norwell, MA (2003).

[7] T. Ide and K. Inoue: Knowledge discovery from heterogeneous dynamic systems using change-point correlations, *Proc. 2005 SIAM International Conference on Data Mining* (*SDM 05*), pp.571-576 (2005).

[8] 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 conference of the Institute of Systems, Control and Information Engineers* (*SCI'09*), pp.409-410, (2009) [in Japanese].