

Temporal Segmentation and Recognition of Body Motion Data Based on Inter-Limb Correlation Analysis

Toru NAKATA

Digital Human Research Center, AIST, Japan. toru-nakata@aist.go.jp

Abstract—A method for segmentation and recognition of human body behavior data is proposed. Recognition of human body movements is getting larger interests in robotic research field, since robots must recognize human behavior in order to interact with human in the real world. In addition, there is demand for quantitative methods to analyze human body movements, since human body movements can be used as models of robot behaviors. The author proposes a scheme for human behavior recognition based on two process steps: analysis of movement correlations among limbs and temporal segmentation of motion data. Inter-limb movement correlations are widely observed in various behaviors and well represent contents of behavior, so it will be a universal feature value for general behavior. Observing changes of inter-limb correlations, we can segment motion capture data into temporal fragment of action units. Using this segmentation technique in an experiment, the system succeeded recognizing various types of human behavior efficiently.

I. INTRODUCTION

A. Motivation

Today, motion capture technology allows us measuring human body motions precisely. We are facing an open problem: how can intelligent systems use motion capture data. Automatic recognition of human behavior is one of the common challenges over researches about human motion analysis.

Motion capture data, however, are not easy to recognize in general. The data usually have large amount of dimensions to record XYZ trajectory of major parts of a human body. High dimensionality makes recognition difficult.

In addition to this, inconsistency of measuring protocol over engineers disturbs compatible use of motion capture data. Each engineer is selecting number and positions of the markers on human bodies independently. Even though a motion looks similar for human eyes, form and content of the captured data may be greatly different.

Here is my question. Is such large amount of data really required for motion recognition? Do we distinguish human activities by calculating XYZ trajectories of 10 or more points of the body? Must the measuring points be uniformed?

According to result of Johansson's famous experiment, our cognitive systems require very small and disunion motion data [1]. Details of human body motion are not important to distinguish human behavior. Also, we can recognize so quickly. It means the calculation for recognition is not complicated.

The behavior recognition mechanism, therefore, must have a process for reducing data dimensions that consists of simple mathematical structure.

I propose a recognition technique that greatly reduces size of motion capture data in respect to space and time. Details of data should be merged into small number of features that are able to represent human body behavior. Regarding time, dense time-series of motion capture data ought to be adequately segmented into blocks of actions, so that data will be sparse yet efficient to describe body motion.

B. Related Works

Many researchers attach importance to temporal segmentation of motion capture data of human body action. Segmentation makes the data easy to use for computers in purpose of edit, recognition, and motion generation.

Segmentation is a process that finds discontinuity in the time-series of data. Data chunks between the gaps are considered as primitive elements of body actions.

In the previous works, there are 4 types of definition about the gap.

First one is that motion stillness means the gap. Several researchers proposed to segment motion capture data at times of body movement with small velocity or kinematical energy [2, 3]. But dynamic range of human body motion often spreads so widely, so it is not easy to find adequate threshold of velocity or energy to define motion stillness.

Second type of the definitions is that accomplishments of certain body poses denote end of action units. Campbell and Bobick [4] segmented dance motion data in respect to pose of foot steps. Since this technique requires definition the key poses, it is difficult to apply for general human motions (not restricted within dances).

Third segmentation policy is that motion data should be segmented by statistics data grouping technique [5-9]. Statistical methods, such as K-mean algorithm or Singular Value Decomposition, have been employed to classify time frame of motion data.

Forth idea is that change of dynamical system stands for the chasm of action [10-12]. They assume that data of motion obey a certain equation for a short time period. Especially dynamical sports movements (such as running) are easy to be formulated with mathematical functions and models, such as Hidden Markov Model and State Space Representation. Many researchers have adopted this way (and often the third idea together) for segmentation and other analysis.

I consider that the third and forth of traditional techniques have weak points as follows. These techniques generally require large number of data samples for their machine learning. In addition, the data samples must have mathematical

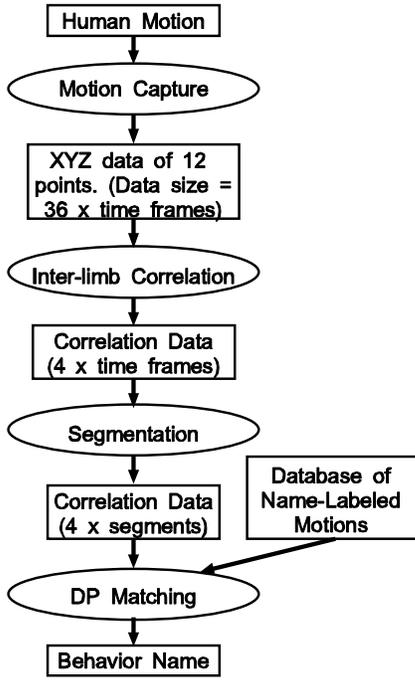


Fig. 1. Process Flow

consistency for formulation of dynamical equations, even though human body motions tend to vary. I think these segmentation policies are kinds of brute force methods demanding large number of good samples and high computing power.

Quest for more efficient and robust methods is still valuable.

II. REDUCTION OF MOTION DATA DIMENSION

Although the human body has huge number of muscles, we cannot control them independently. Most of muscles are controlled in groups [13, 14].

Particularly human 4 limbs are typical units of the packaged control. Extensions and flexions of the muscles in a limb often synchronize in general movements.

Observing human motions more precisely, we can find movement correlations between the 4 limbs. For example, a human moves his right arm and left leg together, when he is walking.

In general, the combination of inter-limb movements is maintained until control of the whole body action changes.

I, therefore, propose that change of the inter-limb movement correlation can be utilized as indicator of action gap. We may segment motion data more properly than traditional techniques, since this method has 2 advantages as follows: 1) this segmentation is independent from amount of velocities or kinematical energy of the body. As this paper shows below, the method can segment various kinds of behavior data from slow movements to active sports movements without tuning the threshold. 2) The method allows minor differences on marker positions of motion capture measurement. The method requires only information about the as-a-whole flexions and extensions of the 4 limbs.

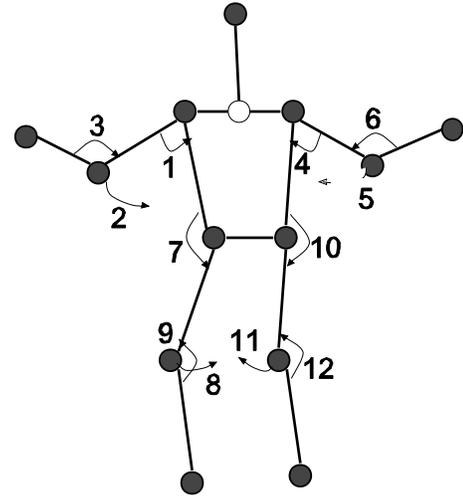


Fig. 2. Major joint angles used for input.

As input for the segmentation system, I selected 12 joint angles of a human body described in Figure 2, which are 4 angles of vertical abductions and 4 angles of horizontal abductions on the shoulders and the hip joints, and 4 angles of flexions on the elbows and the knees.

As the first step of the process, the system measures the 12 joint angles and then calculates degree of flexion of each limb. For instance, degree of flexion of the right arm is formulated as following:

$$Flexion_{RightArm} = \dot{\theta}_1 + \dot{\theta}_2 + \dot{\theta}_3, \quad (1)$$

$$Flexion_{LeftArm} = \dot{\theta}_4 + \dot{\theta}_5 + \dot{\theta}_6, \quad (2)$$

$$Flexion_{RightLeg} = \dot{\theta}_7 + \dot{\theta}_8 + \dot{\theta}_9, \quad (3)$$

$$Flexion_{LeftLeg} = \dot{\theta}_{10} + \dot{\theta}_{11} + \dot{\theta}_{12}. \quad (4)$$

Then the system computes intensities of the inter-limb movement correlations. In general, mammal body movements have typical 4 patterns of inter-limb correlations.

$$\begin{aligned}
 Corr &= \begin{pmatrix} Corr_1 \\ Corr_2 \\ Corr_3 \\ Corr_4 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} Flex_{RArm} \\ Flex_{LArm} \\ Flex_{RLeg} \\ Flex_{LLeg} \end{pmatrix} \quad (5)
 \end{aligned}$$

The system reduces data size by transforming raw motion capture data (of 39 dimensions) to the inter-limb correlation patterns (of 4 dimensions)¹. Figure 3 shows an example of the inter-limb correlations.

¹There are 2 reasons for using the matrix in Eq.6. Dance researchers concerning Laban Movement Analysis (LMA) consider that the 4 column vectors of the matrix represent typical and usual motion correlations among 4 limbs of vertebrate animals[14]. I do not try to verify such biological argument in this paper, but try to show its effectiveness for motion recognition. In addition, the 4 vectors are special, because they formulate "quaternion group." They often appear as eigen vectors of mechanical vibrations in many structures.

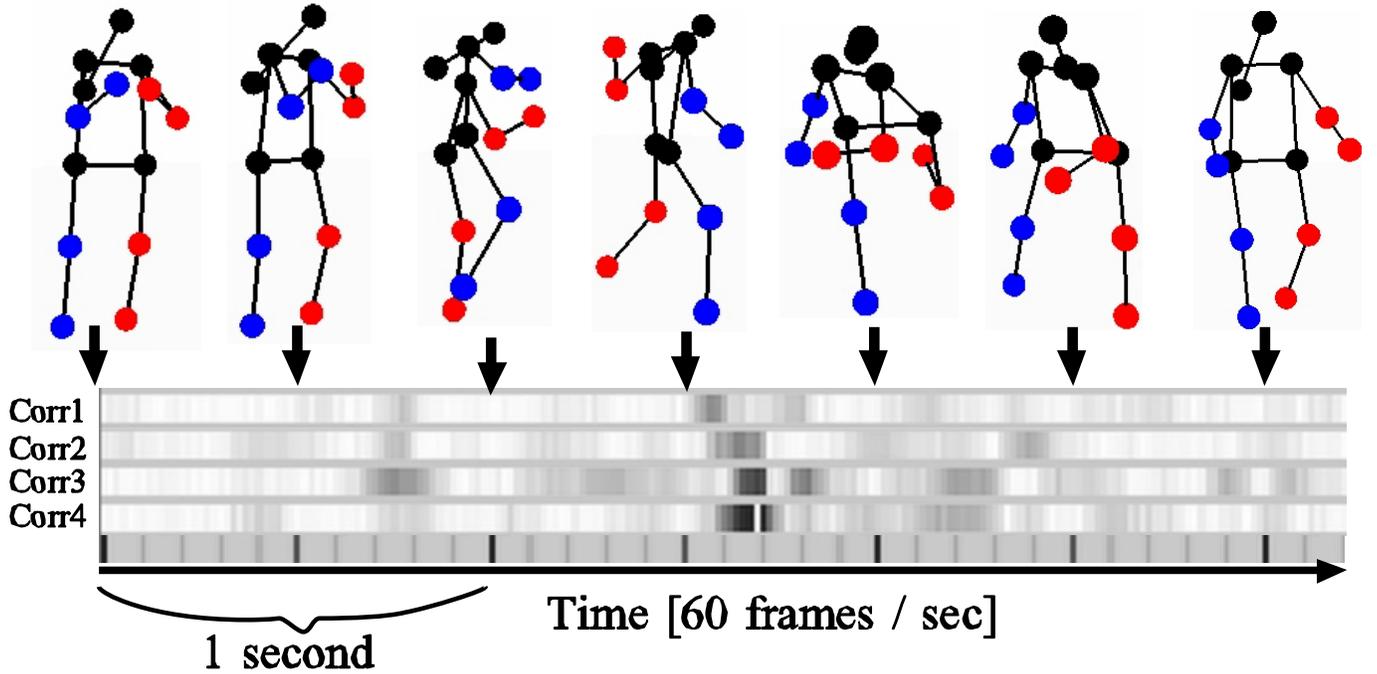


Fig. 3. Linter-limb correlations of a throwing behavior. Darkness of bands indicate absolute values of the correlation intensities.

- 1) Set $t_{Head} = (\text{Frame number of a segment head})$.
- 2) Set $t = t_{Head} + 1$
- 3) Calculate inner product $P = Corr(t_{Head}) \cdot Corr(t)$.
- 4) If $P \geq 0$, go back to step 2.
- 5) Report that t is the head frame of the next segment.
- 6) Find the $Corr$ vector with the largest norm within the segment and report it as the feature vector representing movement of the segment.
- 7) Go back to step 1.

Fig. 4. Segmentation Algorithm

III. SEGMENTING MOVEMENT DATA

The system finds gaps of the motion data by using the feature vector (shown in Eq.6) consisting of intensity values of the inter-limb correlations. Temporal discontinuity of the feature vector indicates that the control of the body motion is drastically changes at its time point.

The system detects gaps of actions under policy described as follows: for any pair of time points in same action segment, inner products of feature vectors between the time points must not be negative value. Figure 5 describes the detail of the segmentation algorithm.

Let see the segmentation results shown in Figure 5 and 6. Those figures show estimated edges of segments. For example, the segmentation results of the walking data well represent foot steps. For active sports movement data such as throwing, the system could extract units of action. (All motion samples were performed by the same person.)

By assuming that body action maintains its character within a segment, the system chooses 1 time points from each segment to represent all data in the segments. In this

experiment, the system selects the time points that bear the maximum norm of the feature vector. Thus, data size shrinks from order of dense sampling frequency to order of sparse segments.

IV. RECOGNIZING BEHAVIOR

Validity of the segmentation must be proven quantitatively. So I employ the segmentation method as a preparation process of behavior recognition.

By the process described above, each motion capture footage are now characterized by sparse time sequences of the inter-limb correlation vectors. Dynamic Programming matching (DP matching) is most suitable to recognize such sequences, since DP matching does not require many samples for learning. (If you have many samples, HMM may achieve better recognition result.)

As a learning signal for the system, I select 11 sample data of different behaviors shown in Figure 3 and 4, namely “walking,” “backward walking,” “lifting,” “sitting on chair,” “sitting of the floor,” “quick standing up,” “slow standing up,” “throwing,” “catching and throwing,” “crawling on the floor,” and “sword play.”

Using DP matching algorithm, the system classifies inputted data into the 11 behavior types by observing difference between inputted data and the 11 templates for learning. I set parameters of DP algorithm as follows: penalty for mismatch is 3, and penalty for stagnation is 1.

For each type of behavior, 2 samples are fed to the system.

Table I shows all of the recognition results. The rate of correct classification was 68.2%. This is quite good result comparing other researches, since most of the researches carried out recognition trials with more few of behavior types.

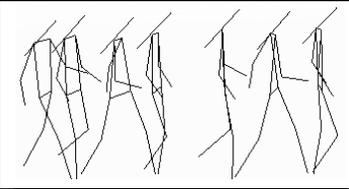
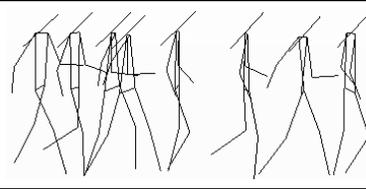
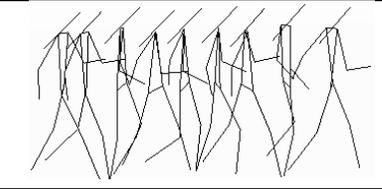
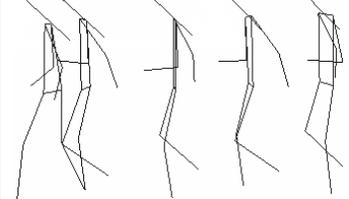
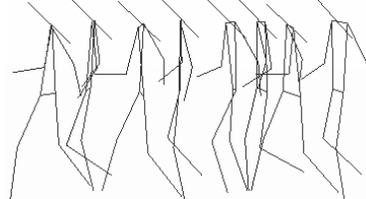
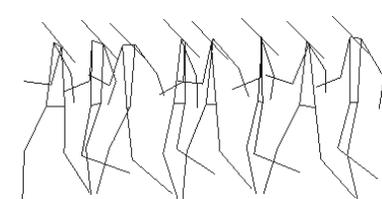
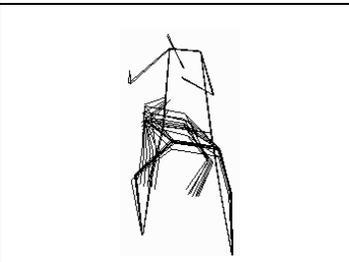
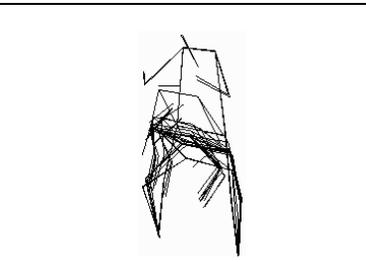
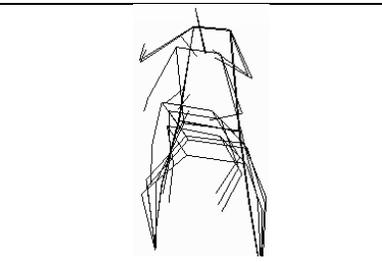
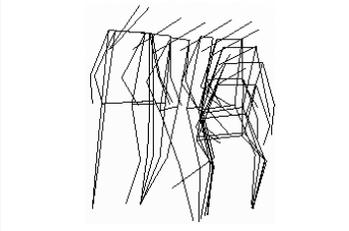
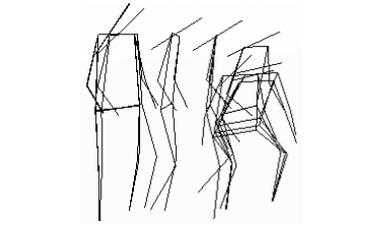
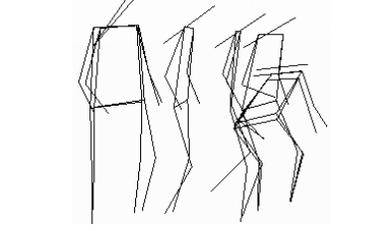
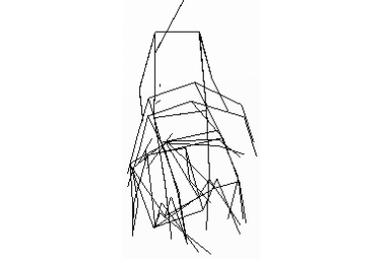
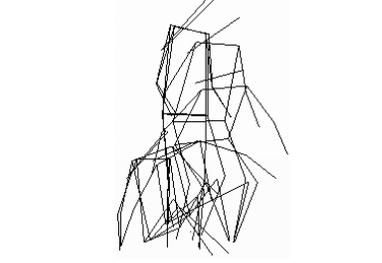
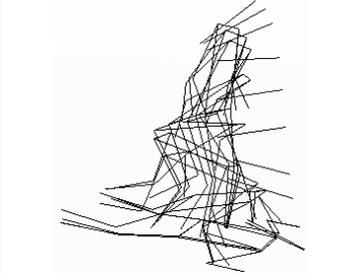
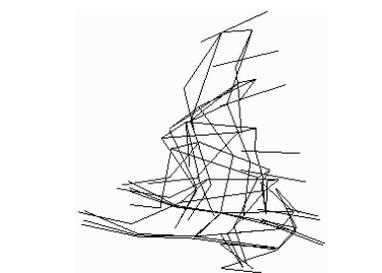
	Sample for Teaching	Samples to be Recognized	
Walking			
Back Walking			
Lifting			
Sitting on Chair			
Sitting on Floor			
Quick Standing up			

Fig. 5. Segmentation Result: Key Frames of Motion Data

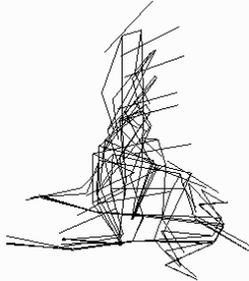
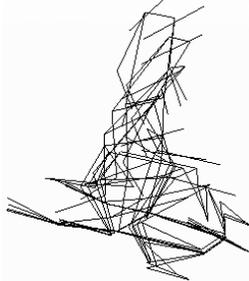
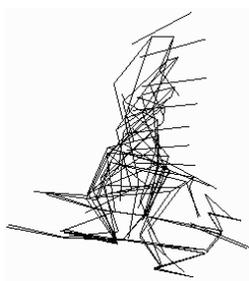
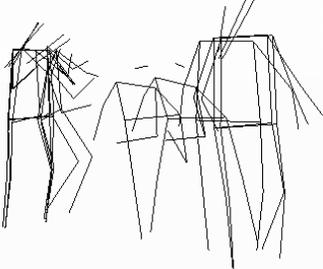
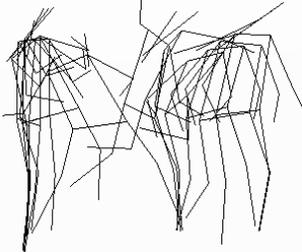
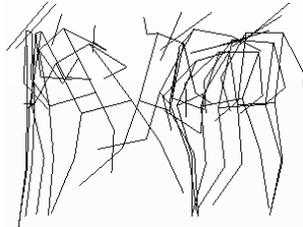
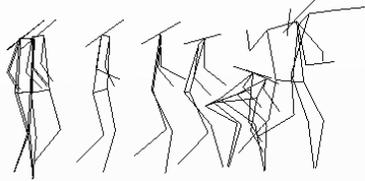
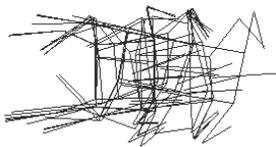
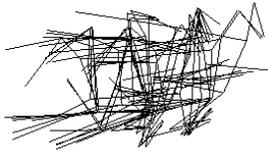
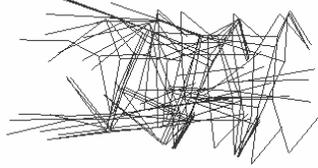
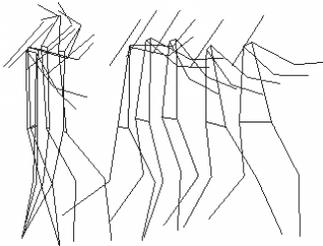
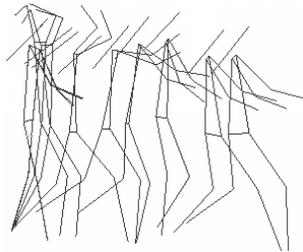
	Sample for Teaching	Samples to be Recognized	
Slow Standing up			
Throwing			
Catching & Throwing			
Crawling			
Sword Play			

Fig. 6. Segmentation Result: Key Frames of Motion Data (Continued)

The result also proves validity of the data size reduction. For comparison, I tried the same process of recognition by inputting velocities of the marker points instead of the feature vector. Even though the data of velocity sequences have large amount of information, the recognition failed more than the segmentation method (Table II).

Inputting raw and dense time series of the major joint angles (without processing temporal segmentation), the recognition got the best result (Table III). In this case, however, inputted data are large.

Considering data size and recognition rate, the segmentation and recognition algorithm using the inter-limb correlations can be said as the most efficient among the 3 processes (Table III).

V. CONCLUSION

This paper proposed an algorithm for segmentation and recognition of human behavior from motion capture data. Apart from details of body motion data, the algorithm observes whole-body-level features, which are movement correlations among the 4 limbs. The correlations allowed the system segmenting the motion footage correctly, and the recognition by DP matching succeeded in high rate with small number of input information.

For future work, I am interested in recognition of multiple behaviors in long footages of motion-capture data. Human behaviors sometimes mix simultaneously and transit from one to another. To process those complex phenomena, the algorithm of the recognition part must be reconsidered.

TABLE I

CONFUSION MATRIX OF RECOGNITION RESULT USING INTER-LIMB CORRELATION. RECOGNITION RATE IS 68.2%.

Behavior	1	2	3	4	5	6	7	8	9	10	11
1) Walking	2										
2) Backward walking	1									1	
3) Lifting			2								
4) Sitting on chair	1			1							
5) Sitting on floor				1				1			
6) Quick standing up						2					
7) Slow standing up							2				
8) Throwing								2			
9) Catching & Throwing				1					1		
10) Crawling			1							1	
11) Sword play											2

TABLE II

CONFUSION MATRIX OF RECOGNITION USING RAW DATA OF THE VELOCITIES. RECOGNITION RATE IS 45.5%.

Behavior	1	2	3	4	5	6	7	8	9	10	11
1) Walking	2										
2) Backward walking				2							
3) Lifting			2								
4) Sitting on chair	1			1							
5) Sitting on floor	1				1						
6) Quick standing up		2									
7) Slow standing up							1	1			
8) Throwing								1	1		
9) Catching & Throwing	1								1		
10) Crawling			1					1			
11) Sword play	1										1

REFERENCES

- [1] Johansson, G., "Visual motion perception," *Scientific American*, Vol.232, pp.76-88, 1975.
- [2] A. Nakazawa, S. Nakaoka, K. Ikeuchi, and K. Yokoi, "Imitating human dance motions through motion structure analysis," *Proc. IEEE/RSJ IROS 2002*, pp.2539-2554, 2002.
- [3] K. Kahol, P. Tripathi, S. Panchanathan, "Automated gesture segmentation from dance sequences," *Proc. 6th IEEE Intl. Conf. Automatic Face and Gesture Recognition*, pp.883-889, 2004.
- [4] L. Campbell, and A. Bobick, "The recognition of human body motion using phase space constraints," *Proc. 5th Intl. Conf. Computer Vision*, pp.624-630, 1995.
- [5] B. Janus, and Y. Nakamura, "Unsupervised probabilistic segmentation of motion data for mimesis modeling," *Proc. IEEE Intl. Conf. Advanced Robotics 2005*, pp.411-417, 2005.
- [6] A. Ali, and J. K. Aggarwal, "Segmentation and recognition of continuous human activity," *Proc. IEEE Workshop on Detection and Recognition of Events in Video*, 2001.
- [7] Y. Rui and P. Anandan, "Segmenting visual actions based on spatio-

TABLE III

CONFUSION MATRIX OF RECOGNITION USING RAW DATA OF THE JOINT ANGLES. RECOGNITION RATE IS 77.3%.

Behavior	1	2	3	4	5	6	7	8	9	10	11
1) Walking	2										
2) Backward walking	2										
3) Lifting			1				1				
4) Sitting on chair				2							
5) Sitting on floor					2						
6) Quick standing up						1				1	
7) Slow standing up							2				
8) Throwing				2							
9) Catching & Throwing				2							
10) Crawling							1		1		
11) Sword play											2

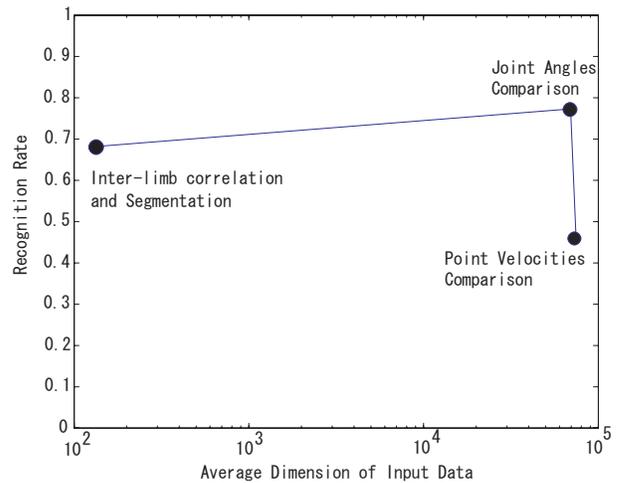


Fig. 7. Average input data size vs. recognition rates.

- temporal motion patterns," *IEEE Conf. Computer Vision and Pattern Recognition*, pp.111-118, 2000.
- [8] T. Mori, Y. Nejigane, M. Shimosaka, Y. Segawa, T. Harada and T. Sato, "Online recognition and segmentation for time-series motion with HMM and conceptual relation of actions," *Proc. IEEE/RSJ IROS 2005*, pp.2856-2863, 2005.
- [9] G. Ma and X. Lin, "Typical Sequences Extraction and Recognition," *European Conference on Computer Vision (ECCV) 2004*, LNCS 3058, pp.60-71, Springer-Verlag, 2004.
- [10] C. Lu, H. Liu, and N. J. Ferrier, "Multidimensional motion segmentation and identification," *Proc. Computer Vision and Pattern Recognition 2000*, pp.629-636, 2000.
- [11] J. Barbic, A. Safonova, J. Y. Pan, C. Flautsos, J. K. Hodins, N. S. Pollard, "Segmenting motion capture data into distinct behaviors," *Graphics Interface*, pp.185-194, 2004.
- [12] C.M. Lu and N. J. Ferrier, "Repetitive Motion Analysis: Segmentation and Event Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2), 258-263, 2004.
- [13] S. R. Robinson and W. P. Sotherman, "Environmental determinants of behaviour in the rate fetus. II The emergence of synchronous movement," *Animal Behaviour* 35, 1625-1662, 1987.
- [14] Bonnie B. Cohen, *Sensing, feeling and action*, Contact Editions, 1993.